

AD-A253 561



2

PUBLICATION OF PAPERS PRESENTED AT URSI 23RD GENERAL ASSEMBLY
AS A SPECIAL SECTION OF RADIO SCIENCE

FINAL REPORT

K. C. YEH

MAY 15, 1992

U.S. ARMY RESEARCH OFFICE

GRANT NUMBER DAAL03-91-M-0149

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
UNIVERSITY OF ILLINOIS
URBANA, IL 61801-2991

APPROVED FOR PUBLIC RELEASE;

DISTRIBUTION UNLIMITED

92 3

92-21132



THE VIEWS, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

| | | | | |
|--|---|--|---|--|
| 1. AGENCY USE ONLY (Leave blank) | | 2. REPORT DATE May 15, 1992 | 3. REPORT TYPE AND DATES COVERED Final 17 Feb 91 - 31 Jan 92 | |
| 4. TITLE AND SUBTITLE Publication of Papers Presented at URSI 23rd General Assembly as a Special Section of Radio Science | | | 5. FUNDING NUMBERS DAAL03-91-M-0149 | |
| 6. AUTHOR(S) K. C. Yeh | | | | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Electrical and Computer Engineering University of Illinois Urbana, IL 61801-2991 | | | 8. PERFORMING ORGANIZATION REPORT NUMBER | |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709-2211 | | | 10. SPONSORING/MONITORING AGENCY REPORT NUMBER ARO 28787.1-GS | |
| 11. SUPPLEMENTARY NOTES The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation. | | | | |
| 12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited. | | | 12b. DISTRIBUTION CODE | |
| 13. ABSTRACT (Maximum 200 words) A symposium entitled "Radio Propagation in the Ionosphere and Magnetosphere: Theory and Application" was convened at the 23rd General Assembly of URSI, August 28 through September 5, 1990. Of the 49 papers presented at this symposium, 13 were selected for publication in Radio Science, Volume 27, Number 2, March-April 1992. These papers cover a broad field of ionospheric and magnetospheric propagation of both experimental and theoretical nature. | | | | |
| 14. SUBJECT TERMS Ionospheric propagation, magnetospheric propagation | | | 15. NUMBER OF PAGES 132 | |
| | | | 16. PRICE CODE | |
| 17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED | 18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED | 19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED | 20. LIMITATION OF ABSTRACT UL | |

DTIC QUALITY INSPECTED 8

| | |
|--------------------|-------------------------------------|
| Accession For | |
| DTIC | <input checked="" type="checkbox"/> |
| DTIC TAB | <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Distribution | |
| By | |
| Distribution/ | |
| Availability Codes | |
| Dist | Avail and/or Special |
| A-1 | |

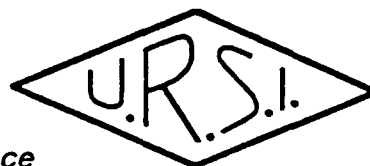
RADIO SCIENCE

Volume 27 Number 2 March-April 1992

*Including Special Section:
Radio Propagation
in the Ionosphere and Magnetosphere:
Theory and Application*



Published by
American Geophysical Union
Cosponsored by
International Union of Radio Science



RADSCAD 27(2)197-368(1992)
ISSN 0048-6604

Foreword

At the 23rd General Assembly (GA) of the International Union of Radio Science (URSI) held in Prague, Czechoslovakia, August 28 through September 5, 1990, a Joint Symposium 2 (JS2), cosponsored by Commissions G and H of URSI, was convened. This symposium was entitled "Radio Propagation in the Ionosphere and Magnetosphere: Theory and Application." The undersigned were appointed as conveners by Commission G Chair, Henry Rishbeth, and Commission H Chair, Hiroshi Matsumoto. For this symposium, altogether 49 papers were received and presented by authors from 12 countries around the world. The richness and diversity of the work marked a milestone of our progress in this field. Many scientific leaders in the radio science community felt that it was perhaps a good opportunity to collect these papers and have them published in one journal as a special section. After a survey of all authors, strong interest had been expressed in publishing them in *Radio Science*. With the strong support of the authors and the Editors of *Radio Science* we proceeded to contact the authors and to process the papers after having received them.

The papers presented at the JS2 of URSI GA covered a vast field. In terms of geographic regions, propagation paths under study took place in high latitudes where the auroral ionosphere, under the perpetual influence of penetrating convective electric fields of magnetospheric origin, was often very disturbed, or in equatorial latitudes where the ionosphere was subject to equatorial dynamics and plasma instabilities, or over long paths traversing many different regions of the globe. In terms of wave frequencies the waves could be in the VLF band propagating in the Earth-ionosphere waveguide mode; they could also be ion cyclotron waves and whistler waves which were slowly propagating and highly dispersive; they could also be in the HF to SHF bands where effects produced by the inhomogeneous and anisotropic nature of the ionosphere were manifested. In terms of ionospheric phenomena the papers contained studies of traveling ionospheric disturbances and propagation effects through equatorial bubbles, chemically modified ionospheres, and ionospheric irregularities created by plasma instabilities. In order to handle and describe propagation in such a complex and rich environment, mathematical techniques used by several branches of physics had to be adapted. These methods were used to study wave propagation and radiation in a medium that could be random, inhomogeneous, anisotropic, dispersive, and nonlinear.

The papers in this section are roughly grouped into theory, HF propagation, and VLF and ELF propagation. In putting this special section together we have been encouraged by Alan T. Waterman, Jr., the immediate past Editor, and David C. Chang, the current Editor. Numerous reviewers have been extremely helpful and cooperative in providing constructive comments, sometimes with very demanding and pressed deadlines. Much editing on some manuscripts received at Illinois was carried out ably by Belva Edwards. Partial support for the publication of this special section was provided by the U.S. Army Research Office. To all of them, we acknowledge gratefully with thanks and pleasure.

K. C. Yeh
Iwane Kimura
Guest Editors

Copyright 1992 by the American Geophysical Union.

Paper number 91RS03050.

0048-6604/92/91RS-03050\$02.00

The generalization of Rytov's method to the case of inhomogeneous media and HF propagation and scattering in the ionosphere

N. N. Zernov, V. E. Gherm, N. Yu. Zaalov, and A. V. Nikitin

Institute of Physics, University of St. Petersburg, Ulyanovskaya, Petrodvorets, Russia

(Received January 23, 1991; revised October 14, 1991; accepted October 15, 1991.)

Rytov's method is generalized for the case when the regularly inhomogeneous ionosphere contains local irregularities of deterministical and statistical nature. An integral representation of the point source field is constructed which takes into account the multibeam effects including strong interference of the beams field, i.e., caustics and foci formation and so on. By means of the representation developed, some problems of HF wave propagation are solved. Among these is the diffraction of a point source field on an ambipolar diffusing local inhomogeneity of the ionosphere. The calculations of the power spectra of the fluctuations of the phase and amplitude of a single-hop path field due to ionospheric electron density irregularities are also presented. The calculations are compared with the results of experimental investigations.

1. INTRODUCTION

The necessity of the HF field focusing description stimulated the development of the point source field asymptotic integral representations. There are now some well-known methods to accomplish this. The most simple method is the Fourier integral for the point source field in the plane-stratified medium, still used by Brekhovskikh [1960] and Budden [1961]. The mathematical aspects of the problem were widely investigated by some mathematicians like Maslov [1965], Ludwig [1966], and others. In the applied form, Maslov's method is described by Kravtsov [1968] and Ziolkowski and Deschamps [1984]. The integral representation founded on geometrical optics has been developed by Orlov [1972]. His results have been used by Tinin [1983] for the case of HF scattering in a fluctuating ionosphere. The papers mentioned do not take into account the diffraction effects in the case of inhomogeneities of the spatial scale not large in comparison with Fresnel zone size.

The well-known phase screen method and the classic Rytov's method give the opportunity to describe diffraction, but there are a lot of difficulties to take regular refraction into account in the frames of these methods and use them for a point source field. Regular refraction and diffraction effects may be described by the phase screen-diffraction layer

method [Wagen and Yeh, 1986, 1989] for the case of arbitrary plane wave incidence on the ionosphere, when random phase changes and diffraction are taken into account separately and alternatively.

The present paper is devoted to the generalization of Rytov's method, giving the description of the point source field in a medium with regular refraction, containing local irregularities, particularly of not very large spatial scales, when the diffraction effects are significant. In section 2, Rytov's method is generalized to the form of an integral representation, describing multibeam effects, including field focusing on caustics and others. For each partial wave of the integral representation the diffraction effects on local irregularities with nonzero wave parameter are taken into account by means of a complex phase. Section 3 gives an example of the application of the generalized method to the case of a local disturbance in the ionosphere. The HF propagation in the presence of random inhomogeneities is considered in section 4, where both theoretical and experimental results are presented and compared. Section 5 contains the conclusions.

2. THE GENERALIZATION OF RYTOV'S METHOD

In the isotropic approximation for HF waves, let us consider the scalar equation,

$$\nabla^2 U + k^2[\epsilon_0(z) + \epsilon(r, t)]U = \delta(r), \quad (1)$$

where k is a wave number, δ is Dirac's function, $\epsilon_0(z)$ represents the plane-layered undisturbed ion-

Copyright 1992 by the American Geophysical Union.

Paper number 91RS02920.
0048-6604/92/91RS-02920\$08.00

osphere model, and ϵ represents a regular or random ionospheric irregularity. In the case of stochastic problems the mean value $\langle \epsilon(r, t) \rangle$ is supposed to be zero.

The characteristic spatial and time scales of the irregularities l_ϵ and T_ϵ are assumed to satisfy the conditions $kl_\epsilon \gg 1$ and $\omega T_\epsilon \gg 1$, where ω is the circular frequency.

Let us begin with the case of an undisturbed ionosphere. It is well known [Brekhovskikh, 1960; Budden, 1961] that the asymptotic solution of the reduced ($\epsilon = 0$) equation

$$\nabla^2 U_0 + k^2 \epsilon_0(z) U_0 = \delta(r) \quad (2)$$

can be written in the form of a Fourier integral

$$U_0(r) = \frac{i}{4\pi} \int_{-\infty}^{+\infty} a_0(r, \alpha) d\alpha \quad (3)$$

with

$$a_0(r, \alpha) = [\epsilon_0(z) - \alpha^2]^{-1/4} [\epsilon_0(0) - \alpha^2]^{-1/4} \cdot \exp [ik\alpha x + ikL_{1,2}(z, \alpha)], \quad (4)$$

where the dependence of the variable z is constructed in WKB approximation and the phase function $L_{1,2}$ has the form

$$L_1(z, \alpha) = \int_0^z [\epsilon_0(z) - \alpha^2]^{1/2} dz, \quad (5)$$

if there is no reflection from the ionosphere, and

$$L_2(z, \alpha) = \int_0^z [\epsilon_0(z) - \alpha^2]^{1/2} dz + \int_z^{\bar{z}} [\epsilon_0(z) - \alpha^2]^{1/2} dz - \frac{\pi}{2k} \quad (6)$$

with the phase shift $(-\pi/2k)$ for the wave, reflected from the ionosphere on the level \bar{z} , determined by the equality $\epsilon_0(\bar{z}) - \alpha^2 = 0$. If the medium is homogeneous, expressions (3) and (4) with L_1 from (5) give the point source field expansion by plane waves.

Despite of the formal invalidity of WKB partial waves (4) near their turning points (caustics $z = \bar{z}$), the integral (3) is known [Maslov, 1965] to be the uniform asymptotic integral representation of the point source field in a plane-stratified slowly varying media. For further calculations of the field $U(r)$ it is necessary to investigate the saddle points of the

integrand in (3) and consider the cases of the separated (for independent rays) and degenerated (for caustics and so on) saddle points.

Let us return now to the full equation (1). We construct its solution in the same form as (3):

$$U(r, t) = \frac{i}{4\pi} \int_{-\infty}^{+\infty} a(r, t, \alpha) d\alpha \quad (7)$$

In the case of the small $\epsilon(r, t)$ the partial waves (4) are slightly disturbed, and we represent the unknown function $a(r, t, \alpha)$ as a perturbation series

$$a(r, t, \alpha) = a_0(r, \alpha) + a_1(r, t, \alpha) + \dots \quad (8)$$

The second term of the series (8), $a_1(r, t, \alpha)$, is the solution of the first-order equation,

$$\nabla^2 a_1 + k^2 \epsilon_0(z) a_1 = -k^2 \epsilon(r, t) a_0 \quad (9)$$

which means it is given by the expression

$$a_1(r, t, \alpha) = -k^2 \int U_0(r, r') \epsilon(r', t) a_0(r', \alpha) dr', \quad (10)$$

where $U_0(r, r')$ satisfies (2) with the right-hand side of the form $\delta(r - r')$.

The next step is to represent the sum $a_0 + a_1$ in (8) as

$$a_0(r, t) + a_1(r, t, \alpha) = a_0(r, t) \exp [\psi_1(r, t, \alpha)], \quad (11)$$

$$\psi_1(r, t, \alpha) = \frac{a_1(r, t, \alpha)}{a_0(r, t)}$$

Therefore one can finally write the solution of the initial equation (1) as

$$U(r, t) = \frac{i}{4\pi} \int_{-\infty}^{+\infty} a_0(r, t) \exp [\psi_1(r, t, \alpha)] d\alpha. \quad (12)$$

The representation of the sum (11) in exponential form corresponds to the transition to Rytov's approximation for the separate "plane" wave propagation description [Tatarskii, 1961]. The function $\psi_1(r, t, \alpha)$ can be shown to satisfy the first approximation equation for the complex phase.

The expression (12) is the ultimate result representing Rytov's method generalization for the case of a point source in an inhomogeneous medium. It includes multibeam effects and diffraction.

Further developments follow from (12). One can calculate the integral by some asymptotic method or numerically in the case of regular inhomogeneities. On the other hand, it leads to a statistical

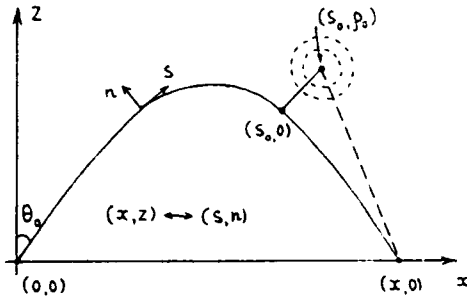


Fig. 1. The geometry and the coordinate system in the problem of HF propagation in the presence of a local inhomogeneity, evolving in accordance to the ambipolar diffusion law: (s_0, p_0) , coordinates of the center of disturbance; s , coordinate along the ray; n , normal to the ray.

characterization of the field in the stochastic problems. These two situations will now be discussed.

3. HF PROPAGATION IN THE IONOSPHERE IN THE PRESENCE OF A LOCAL INHOMOGENEITY

Among the problems of the first kind let us consider the case of a one-hop path, when the transmitter and the receiver are on the Earth's surface and there is a regular (nonrandom) inhomogeneity in the ionosphere. For the sake of simplicity we restrict our consideration to the two-dimensional case when ϵ is a function of two spatial variables and time. As shown in the works by Zernov [1984, 1990], it is suitable to calculate integral represented by formulae (4) and (10)–(12) in the special coordinate system (s, n) (Figure 1), where for each α , s is the variable along the curved beam in the undisturbed ionosphere, beginning at the plane wave front containing the source and passing through the observation point. The variable n is perpendicular to the curved beam at the arbitrary point (x, z) . In these coordinates the complex phase can be calculated and is equal to [Zernov, 1990]:

$$\psi_1(x, 0, \alpha) = -2^{-3/2} \pi^{-1/2} k^{3/2} e^{-i\pi/4}$$

$$\begin{aligned} & \cdot \int \epsilon(s, n) [\epsilon_0(s, 0) - \alpha^2]^{-1/2} M_{\alpha\alpha}^{-1/2}(s, 0) \\ & \cdot \exp \left\{ -\frac{ik\epsilon_0(s, 0)n^2}{2[\epsilon_0(s, 0) - \alpha^2]} M_{\alpha\alpha}^{-1}(s, 0) \right\} \\ & \cdot [1 + np^{-1}(s)] ds dn. \end{aligned} \quad (13)$$

Here the function M is adjusted according to the following rule: $M = L$, if in (4) $L = L_2$, and $M = L_2$ if in (4) $L = L_1$. The function $\rho(s)$ is the radius of curvature of the beam, passing through the observation point.

If the spatial scale l_ϵ is large in comparison to the exponential function scale as a function of n , the integration by n can be performed and the result is the same as the geometrical optics method which gives [Zernov, 1980]:

$$\psi_1^{(0)}(x, 0, \alpha) = \frac{ik}{2} \int \frac{\epsilon(s, 0)}{[\epsilon_0(s, 0)]^{1/2}} ds. \quad (14)$$

In the case of significant diffraction effects with nonzero wave parameter, the expression (13) is the final result and the double integral has to be calculated directly for known models of $\epsilon(s, n)$ or numerically. There exist some model functions of $\epsilon(s, n)$ for which (13) can be calculated.

For example, if

$$\epsilon(s, n, t) = A(t) \exp \{-R^{-2}(t)[(s - s_0)^2 + (n - n_0)^2]\}, \quad (15)$$

where $A(t)$ and $R(t)$ are the time-varying amplitude and spatial scale of the inhomogeneity respectively, the result is

$$\begin{aligned} \psi_1(x, 0, \alpha, t) &= \frac{ik\pi^{1/2}A(t)R(t)}{2[\epsilon_0(s, 0)]^{1/2}} \\ &\cdot (1 - iD)^{-1/2} \exp \left\{ -\frac{\rho_0^2 R^{-2}}{1 - iD} \right\} \end{aligned} \quad (16)$$

with the wave parameter

$$D = \frac{2[\epsilon_0(s, 0) - \alpha^2]M_{\alpha\alpha}(s_0, 0, \alpha)}{kR^2(t)\epsilon_0(s_0, 0)}. \quad (17)$$

Integral (13) can also be calculated for some other kinds of the ionosphere disturbances, for example, in the case of a traveling wave inhomogeneity [Zernov, 1980].

For further calculations of the field (12) it is necessary to find the saddle points satisfying the equation

$$x + \frac{\partial L_2(0, \alpha)}{\partial \alpha} + \frac{1}{ik} \frac{\partial \psi_1(x, 0, \alpha, t)}{\partial \alpha} = 0. \quad (18)$$

Without the last term this equation describes the saddle points for the undisturbed field $U_0(r)$ from (3) and (4). The last term takes in to account the presence of the inhomogeneity. In general, the

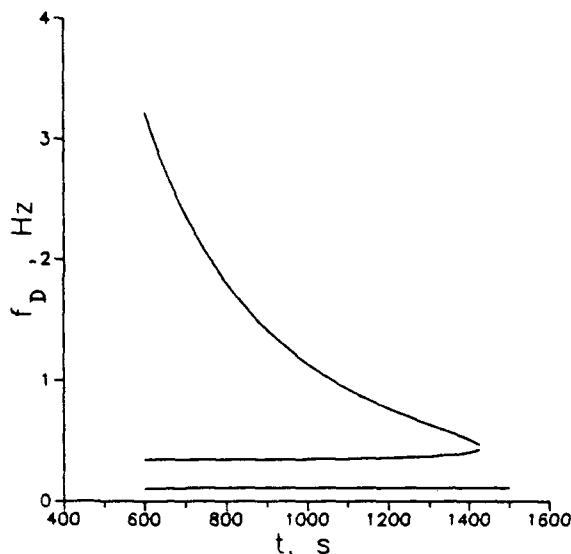


Fig. 2. Temporal dependence of the Doppler frequency shift for every ray.

roots of this equation are complex because of finite value of the wave parameter and have to be calculated numerically. A specially constructed computer program calculates the roots even as a function of time if the parameters of inhomogeneity depend on time and generates the Doppler frequency shift and the intensity of the receiving field.

An example of the calculation of the local disturbance which evolves according to the ambipolar diffusion law is represented in Figures 2 and 3 where, respectively, the temporal dependence of the Doppler frequency shift $f(t)$ and the relative intensity $F(t)$ are plotted for some model parameters. The parameters $R(t)$ and $A(t)$ in (16) are taken in the form

$$R(t) = 2(D_{\perp} t)^{1/2},$$

$$A(t) = -\frac{cQ}{8\pi^{3/2}f_0^2 D_{\perp} D_{\parallel}^{1/2} t^{3/2}},$$

where f_0 is a field frequency, D_{\perp} and D_{\parallel} are the transversal and longitudinal ambipolar diffusion coefficients taken with respect to the Earth magnetic field, Q is the full extra number of charged particles in the disturbance, $C = e^2/\pi m$, where e and m are electron charge and mass. The simulation conditions correspond to a one-hop 1000-km path in a typical summer noon ionosphere.

The presence of a diffraction over the inhomoge-

neity is taken into account by the complex values of the roots. If the roots coincide, the integral (12) is represented by etalon functions with complex arguments. In this case the caustics become complex, i.e., are destroyed.

4. STATISTICAL PROPERTIES OF HF SIGNALS ON ONE-HOP PATHS

4.1. Formulation of the problem and derivation of formulae

In this subsection we derive and analyze the expressions for the phase and intensity correlation functions of the HF field scintillations on one-hop paths in a statistically inhomogeneous ionosphere. The incident field is assumed to be a plane wave of the form (4) propagating with the angle $\theta_0 = \arcsin \alpha_0$ with respect to the normal at the Earth's surface. The one-hop path field in the presence of ionospheric irregularities can be written as

$$E(\mathbf{r}, t) = E_0(\mathbf{r}) \exp \{ \psi_1(\mathbf{r}, t) \}, \quad (19)$$

where ψ_1 is expressed by formula (11). In contrast to (13) it is convenient here to express the phase ψ_1 in rectangular coordinates as

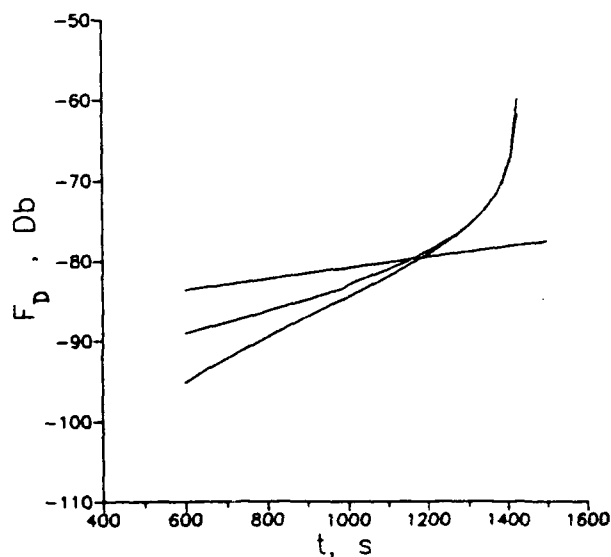


Fig. 3. Temporal dependence of the signal intensity for every ray.

$$\begin{aligned} \psi_1(\mathbf{r}, t) = & \frac{ik^3}{8\pi^2} [\varepsilon_0(z) - \alpha_0^2]^{-1/4} \exp[-ik\alpha_0 x - ikL_2(z, \alpha_0)] \\ & \cdot \int d\mathbf{r}' [\varepsilon_0(z') - \alpha_0^2]^{-1/4} \exp[ik\alpha_0 x' + ikL_{1,2}(z', \alpha_0)] \\ & \cdot \varepsilon(\mathbf{r}', t) \int d\alpha d\beta [\varepsilon_0(z') - \alpha^2 - \beta^2]^{-1/4} \\ & \cdot [\varepsilon_0(z) - \alpha^2 - \beta^2]^{-1/4} \exp[ik\alpha(x - x') \\ & + ik\beta(y - y') + ikM_{1,2}(z, z', \alpha, \beta)]. \end{aligned} \quad (20)$$

Here

$$\begin{aligned} G(\mathbf{r}, \mathbf{r}') = & -\frac{ik}{8\pi^2} \int d\alpha d\beta [\varepsilon_0(z') - \alpha^2 - \beta^2]^{-1/4} \\ & \cdot [\varepsilon_0(z) - \alpha^2 - \beta^2]^{-1/4} \exp[ik\alpha(x - x') + ik\beta(y - y') \\ & + ikM_{1,2}(z, z', \alpha, \beta)] \end{aligned} \quad (21)$$

is the Green function of the undisturbed equation (2) for the case of three spatial dimensions. The functions $M_{1,2}(z, z', \alpha, \beta)$ are the same as in (13), and we will write them for the three-dimensional case,

$$\begin{aligned} M_1(z, z', \alpha, \beta) = & \int_z^{z'} [\varepsilon_0(z) - \alpha^2 - \beta^2]^{1/2} dz \\ & + \int_z^{z'} [\varepsilon_0(z) - \alpha^2 - \beta^2]^{1/2} dz - \frac{\pi}{2k} \end{aligned}$$

if the expression (5) is taken for L and

$$M_2(z, z', \alpha, \beta) = \int_z^{z'} [\varepsilon_0(z) - \alpha^2 - \beta^2]^{1/2} dz$$

if (6) is used.

Having (20) it is easy to calculate the transversal spatial-time spectra

$$\begin{aligned} \hat{\psi}_1(k_x, k_y, z, w) = & (2\pi)^{-3} \int dx dy \psi_1(\mathbf{r}, t) e^{-ik_x x - ik_y y + iwt} \end{aligned} \quad (22)$$

Assuming that $\alpha = \sin \theta_0$, $\beta = 0$ in the slowly varying functions in (20) which corresponds to the assumption that the characteristic spatial scale size of the irregularities l_r is large in comparison with the wavelength of the incidence field λ ; calculating

the integrals over dx, dy in (22) and then over $d\alpha, d\beta$ in (20); and taking into account that

$$\begin{aligned} (2\pi)^{-3} \int dx' dy' dt \varepsilon(\mathbf{r}', t) \exp(-ik_x x' - ik_y y' + iwt) \\ = \hat{\varepsilon}(k_x, k_y, z, w) \end{aligned}$$

is the transversal spatial-time spectrum of the irregularities; we have

$$\begin{aligned} \hat{\psi}_1(k_x, k_y, z, w) = & \frac{ik}{2} \int dz' [\varepsilon_0(z') - \alpha_0^2]^{-1/2} \\ & \cdot \hat{\varepsilon}(k_x, k_y, z', w) \exp \left[ikM_{1,2} \left(z, z', \alpha_0 + \frac{k_x}{k}, \frac{k_y}{k} \right) \right. \\ & \left. + ikL_{1,2}(z', \alpha_0) - ikL_2(z, \alpha_0) \right] \end{aligned} \quad (23)$$

One can obtain transversal spatial-time spectra of the phase and intensity scintillations by extracting the real and imaginary parts of the complex phase (23) and then constructing their second-order moments. Performing some transformations and assuming statistical homogeneity of the irregularities, we obtain the expressions for the phase and amplitude spectra $\hat{B}_{s,x}(k_x, k_y, z, w)$

$$\begin{aligned} \hat{B}_{s,x}(k_x, k_y, z, w) = & \frac{k^2}{4} \int d\xi d\zeta \left[\varepsilon_0 \left(\zeta - \frac{\xi}{2} \right) - \alpha_0^2 \right]^{-1/2} \\ & \cdot \left[\varepsilon_0 \left(\zeta - \frac{\xi}{2} \right) - \alpha_0^2 \right]^{-1/2} \hat{B}_e(k_x, k_y, \xi, w) \\ & \cdot \left\{ 1 \pm \cos \left[\frac{k_x^2 + k_y^2}{2} \int_z^{z'} \frac{d\tau}{[\varepsilon_0(\tau) - \alpha_0^2]^{1/2}} \right] \right. \\ & \cdot \cos \left[\frac{k_x^2 + k_y^2}{k} \int_z^{z'} \frac{d\tau}{[\varepsilon_0(\tau) - \alpha_0^2]^{1/2}} \right] \\ & + \cos \left[2\alpha_0 k_x \int_z^{z'} \frac{d\tau}{[\varepsilon_0(\tau) - \alpha_0^2]^{1/2}} \right] \\ & \cdot \left(\cos \left[\frac{k_x^2 + k_y^2}{k} \int_z^{z'} \frac{d\tau}{[\varepsilon_0(\tau) - \alpha_0^2]^{1/2}} \right] \right. \\ & \left. \left. \pm \cos \left[\frac{k_x^2 + k_y^2}{k} \int_z^{z'} \frac{d\tau}{[\varepsilon_0(\tau) - \alpha_0^2]^{1/2}} \right] \right) \right\}. \end{aligned} \quad (24)$$

where signs "plus" and "minus" correspond to the phase and amplitude spectra, respectively, and \tilde{B}_ϵ is the dielectric constant fluctuation spectrum.

Next we consider an isotropic power law spectrum model for the irregularities. In particular, the spectrum

$$B_\epsilon(\kappa) = \frac{C_\epsilon^2}{(1 + |\kappa|^2/\kappa_0^2)^{p/2}} \quad (25)$$

where C is a normalization constant, $k_0 = 2\pi/\Lambda_\epsilon$, where Λ_ϵ is an outer scale size, and p is a spectral index.

To construct a spatial-time irregularity model, we impose a hypothesis of a "frozen drift" which assumes that the irregularities move with constant speed without changing their profiles in the (x, y) plane. This means we can write

$$\tilde{B}_\epsilon(k_x, k_y, \xi, w) = \delta(w - k_x v_x - k_y v_y) \cdot \int \frac{C_\epsilon^2 \exp(-ik_z \xi) dk_z}{[1 + (k_x^2 + k_y^2 + k_z^2)/\kappa_0^2]^{p/2}} \quad (26)$$

where v_x and v_y are the drift velocity components along the x and y axes, respectively.

To proceed with the evaluation of B_s and B_χ , it is necessary to assume a model for the undisturbed ionospheric profile $\epsilon_0(z)$. For the sake of simplicity we will use the linear model

$$\epsilon_0(z) = 1 - (z/H) \quad (27)$$

where H is the layer height. It should be noted that the frequency spectra of the phase and intensity scintillations, which are measured in our experiments, are connected with the spatial-time frequency spectra (24) by the relation

$$B_{s,\chi}(z, w) = \int dk_x dk_y \tilde{B}_{s,\chi}(k_x, k_y, z, w) \quad (28)$$

Substituting (26) into (24), (24) into (28), and performing the possible integrations, we obtain after some transformations the final expression

$$B_{s,\chi}(w) = \frac{\pi^{1/2} k^2 C_\epsilon^2 A_1^2}{\Gamma(p/2) 2^{(p-5)/2} V (1 - \alpha_0^2)} \cdot \int_0^{\pi/4} \frac{d\varphi}{(\cos 2\varphi)^{1/2}} \int_0^{[2^{1/2} \sin(\varphi + \pi/4)]^{-1}} dz$$

$$\cdot \int_0^\infty dt \frac{F_{p-1} \frac{(2rA_1 \sin \varphi)(1 + W^2 + t^2)}{2}}{(1 + W^2 + t^2)^{(p-1)/2}} \cdot \left\{ 1 \pm \cos [2A_1 A_2 (W^2 + t^2)] \cos [2A_1 A_2 (W^2 + t^2)] \cdot (r \cos \varphi)^{1/2} + \cos \left[\frac{4A_1 \alpha_0}{(1 - \alpha_0^2)^{1/2}} W \cos \psi (r \cos \varphi)^{1/2} \right] \cdot \cos \left[\frac{4A_1 \alpha_0}{(1 - \alpha_0^2)^{1/2}} t \sin \psi (r \cos \varphi)^{1/2} \right] \cdot [\cos (2A_1 A_2 (W^2 + t^2))] \right. \\ \left. \pm \cos [2A_1 A_2 (W^2 + t^2) r \cos \varphi]^{1/2} \right\}, \quad (29)$$

where we have introduced the dimensionless parameters $A_1 = k_0 H (1 - \alpha_0^2)$, $A_2 = (k_0/k)(1 - \alpha_0^2)^{-1/2}$, $W = w/k_0 v$, and $F_\nu(x) = x^\nu K_\nu(x)$, $K_\nu(x)$ being the Macdonald's function of the order of ν . This formula generalizes the expression obtained by *Gherm et al.* [1990] to the case of oblique propagation.

Note that formula (29) includes the WKB approximation as a particular case. The WKB expression for B_s can be obtained from (29) supposing $A_1 A_2 = 0$. In this case the intensity fluctuation spectrum is equal to zero.

To evaluate a triple integral (29), the following Monte Carlo procedure is used. The random three-dimensional vector with components $\{\varphi, r, t\}$ is stochastically generated, each component being distributed in accordance with its own distribution function. Thus the distribution function of the variable t is chosen in the form

$$\sigma_t(t) = \frac{2}{\pi} \arctan \frac{t}{(1 + w^2)^{1/2}},$$

which provides a generation of the random quantity t in the range $[0, \infty]$. The value of t is calculated by the formula

$$t = (1 + w^2)^{1/2} \tan \left(\frac{\pi}{2} \xi_t \right),$$

where ξ_t is a uniformly distributed random number from $[0, 1]$. The distribution function of the variable φ is

$$\sigma_{\varphi}(\varphi) = 1 - [2|\sin(\varphi - \pi/4)|]^{1/2}$$

and for φ we have

$$\varphi = \frac{\pi}{4} - \arcsin\left(\frac{\xi_{\varphi}^2}{2}\right),$$

where ξ_{φ} is a uniformly distributed random number from $[0, 1]$. Such a choice of a $\sigma_{\varphi}(\varphi)$ removes the problem of the singularity when $\varphi = \pi/4$. The results of this evaluation and a comparison with the experimental results are given in the following sections.

4.2. Analysis of the results

First, we put $\theta_0 = 0$ which corresponds to vertical propagation and concentrate on the scintillation spectra and their dependence on the values of the parameters A_1 and A_2 . Some results of the numerical evaluation are represented in Figures 4–6. In all these figures, curves $B_s(w)$ and $B_{\chi}(w)$ are plotted together with the phase spectrum curve evaluated in the WKB approximation. A double logarithmic scale is used.

It is clear from the results of the numerical evaluation that there are some common features in all figures. Namely, in the low-frequency range the

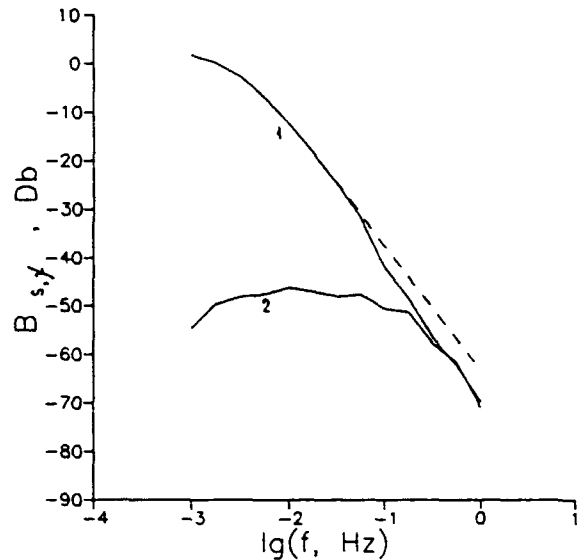


Fig. 5. Calculated phase and intensity scintillation spectra for vertical sounding: $A_1 = 6.28$; $A_2 = 0.0001$; $p = 3.7$. Notation is the same as in Figure 4.

$B_s(w)$ curves evaluated by Rytov's and WKB methods coincide. In contrast, the $B_{\chi}(w)$ curve is situated at much lower values than the $B_s(w)$ curve. This result illustrates the validity of the WKB approximation for the evaluation of the phase scin-

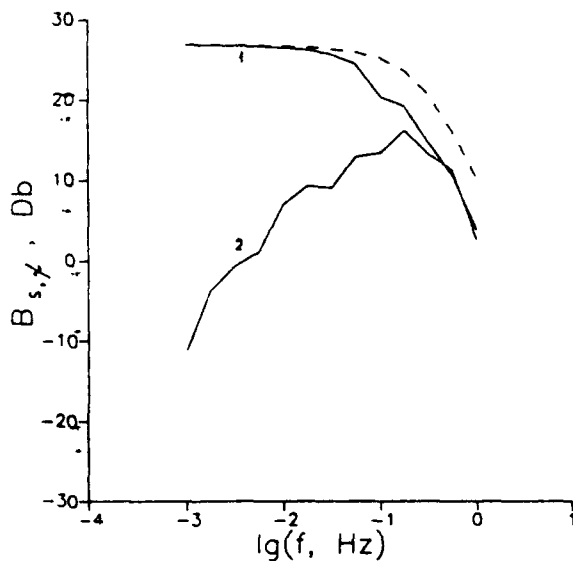


Fig. 4. Calculated phase and intensity scintillation spectra for vertical sounding: $A_1 = 628$; $A_2 = 0.01$; $p = 3.7$; 1. phase spectrum; 2. intensity spectrum; dashed curve, phase spectrum in the WKB approximation.

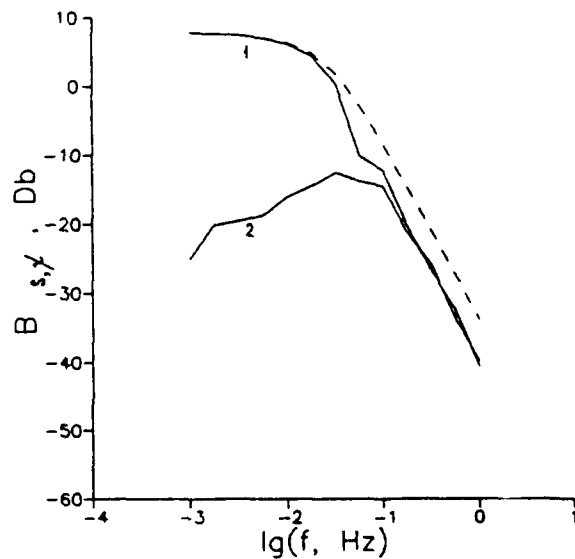


Fig. 6. Calculated phase and intensity scintillation spectra for vertical propagation. Parameters: $H = 150$ km; $\lambda = 30$ m; $\Lambda_e = 10$ km; $c = 300$ m/s; $p = 3.7$; $\sin \theta_0 = 0$; $\psi = 0$.

tillation spectrum in the low-frequency range. It should be noted that phase spectrum when $\omega \rightarrow 0$ is a horizontal line.

When the frequency ω increases, the $B_s(\omega)$ curves evaluated by Rytov's and the WKB methods separate. In contrast, the curves $B_s(\omega)$ and $B_x(\omega)$ approach one another as the frequency increases and then merge. The WKB approximation is not valid in this case; the indicated behavior reflects this fact. It should be noted that the WKB approximation is not valid in all situations when the frequency is sufficiently high.

The high-frequency asymptote of the phase and intensity spectra, as one can prove from (29), is a power law function of frequency of the form

$$B_{s,x}(\omega) \sim \omega^{-(p+1)}. \quad (30)$$

This asymptote is represented by the straight lines in the high-frequency regime of the double logarithmic scale given in Figures 4 and 5.

Analyzing the numerical evaluations of (29) for several different values of A_1 and A_2 , we can ascertain the dependence of the frequency ω_m , where the $B_s(\omega)$ and $B_x(\omega)$ curves merge, on the values of the parameters A_1 and A_2 . In this manner we have determined this dependence to be

$$\omega_m^2 = \frac{0.63}{A_1 A_2} \quad (31)$$

Using the expressions for A_1 , A_2 , and W , we obtain the drift speed $v = (v_x^2 + v_y^2)^{1/2}$

$$V \approx 3.2(\lambda H)^{1/2} f_m (1 - \alpha_0^2)^{1/4}, \quad (32)$$

where dimensioned merging frequency $f_m = \omega_m/2\pi$.

The field wavelength and the ionosphere layer height H are known for every given experiment. Thus (32) allows us to estimate the drift velocity v from the experimentally determined values of the frequency f_m .

So Figures 4–6 have many common features, but they are different in details. Namely, the difference between values of B_s in its maximum when $\omega \rightarrow 0$ and in the point where it merges with $B_x(\omega)$ is about 15 dB in Figure 4, while in Figure 5 this difference reaches 60 dB. This variation is connected with different values of parameters A_1 and A_2 : in Figure 4, $A_1 = 628$ and $A_2 = 0.01$, while in Figure 5, $A_1 = 6.28$ and $A_2 = 0.0001$. If we fix the values of λ , H , v , and p , we find that the only free parameter

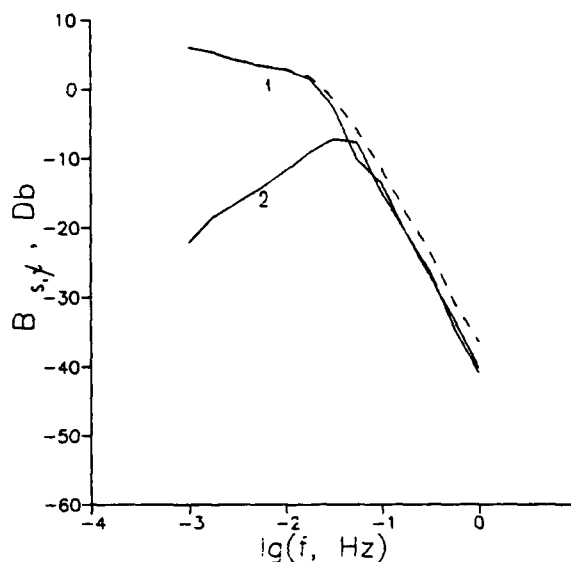


Fig. 7. Calculated phase and intensity scintillation spectra for oblique propagation: $\sin \theta_0 = 0.3$, $\psi = 0$. Other parameters are the same as in the Figure 6.

remaining in the expressions for $B_s(\omega)$ and $B_x(\omega)$ is Λ_ϵ , so its value defines the details of disposition of B_s and B_x curves in the low-frequency range. Therefore it is possible to estimate the outer scale size from the condition of the best coincidence between the theoretical and experimental curves.

For the case of oblique propagation we investigated the influence of the ray incidence angle θ_0 and the direction of the frozen drift of irregularities ψ on the phase and intensity scintillation spectra shapes. A series of calculations with different θ_0 , ψ has been performed to study this behavior. It is found that the character of phase and intensity spectra does not depend significantly on the value of θ_0 but only slightly (Figure 7). As to the drift direction influence, it was found that the drift direction does not practically affect on the shape of the calculated spectra.

The analysis of a great number of calculations revealed that the features of the scintillation spectra are stable in a wide range of variations of the parameters θ_0 , ψ .

4.3. Experimental technique and results

Our experimental equipment provides for the measurement of the HF signal Doppler shift and intensity. The output of the receiver is recorded in

real time by an analog-to-digital converter (ADC) whose sampling rate can vary from 0.5 s^{-1} to 256 s^{-1} . These data are then subjected to a sliding fast Fourier transform (FFT). The FFT frequency resolution depends on the spectral analysis parameters and can vary in a wide range. In our experiments the typical resolution is $\sim 0.03 \text{ Hz}$.

The moment estimates of the Doppler shift F_D and the intensity A_D are defined as

$$F_D = \frac{\sum_{k=1}^N f_k |s_k|^2}{\sum_{k=1}^N |s_k|^2}$$

$$A_D = \sum_{k=1}^N |s_k|^2$$

where f_k is the k th Fourier frequency, s_k is the complex Fourier transform, and N is the number of frequencies in the FFT which is equal to the number of samples in the time interval T .

These estimates give two temporal series representing the Doppler shifts $F_D(t)$ and intensities $A_D(t)$. To determine the experimental estimates of the energetic fluctuation spectra we take an FFT of the quantities F_D and A_D . Since there is a known relation between the phase and frequency,

$$F_D = (2\pi)^{-1} \frac{d\varphi}{dt}$$

the phase spectrum can be found from the frequency spectrum by dividing the latter by the frequency.

The scintillation frequency range for which the calculations of the spectra are reliable is restricted from one side by the measurement duration and from the another side by the temporal lag between samples in the F and the A series. With respect to the experimental results given below this range includes periods from 1 to 500 s.

The measurements of the phase and intensity scintillations have been performed on the oblique path between Moscow and Petergof (700 km) and on the quasi-vertical path between Leningrad and Petergof. The frequencies of the transmitters were 9.996 and 4.055 MHz, respectively. The measurement durations were 10 min.

As shown in Figures 8 and 9, the measured spectra have complicated form even after smoothing. Comparing the experimental and calculated

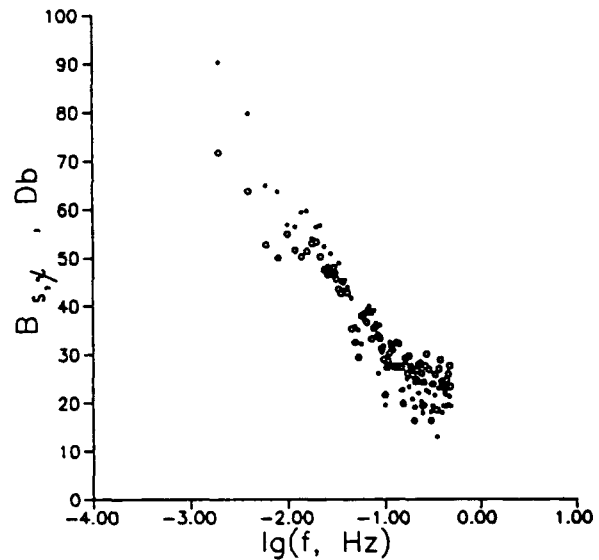


Fig. 8. Experimental phase and intensity spectra: (vertical path) $f_0 = 4.056 \text{ MHz}$.

curves, one can see that the agreement is quite good. The experimental phase and intensity curves merges in the high-frequency range and in the double-logarithmic scale can be approximated by the straight line. Experimental data processing gives a spectral index $p = 3.7 \pm 0.4$, which is in

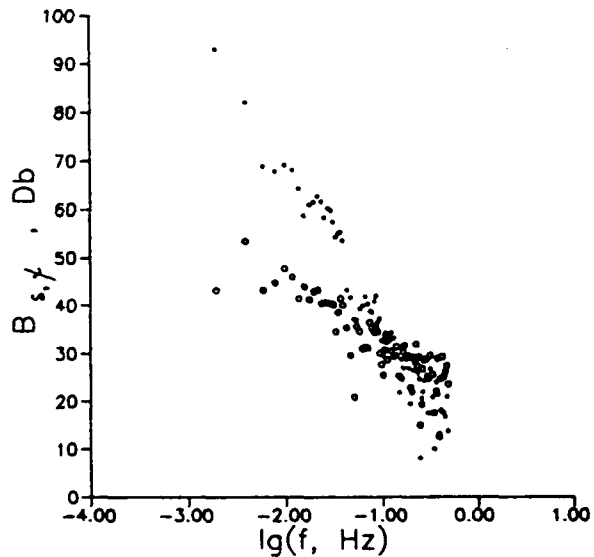


Fig. 9. Experimental phase and intensity spectra: (oblique path) $f_0 = 9.996 \text{ MHz}$.

reasonable agreement with satellite data [Yeh and Liu, 1982].

5. CONCLUSION

Rytov's method has been generalized for the case of a point source in an inhomogeneous media with nonuniform irregularities. An integral representation of the field has been constructed to describe multibeam and diffraction effects. The method presented here has been applied to the cases of propagation in the presence of local disturbances and random inhomogeneities. In the latter case the statistical characteristics predicted by the theory have been compared with those obtained experimentally and the agreement has been found to be quite good.

In conclusion, we note that this generalization of the well-known Rytov's method sufficiently extends the framework of its applicability. The generalized method allows one to solve a variety of real problems which could not be considered within the framework of other methods.

REFERENCES

- Brekhovskich, L. M., *Waves in the Layered Media*, Academic, San Diego, Calif., 1960.
- Budden, H. G., *Radio Waves in the Ionosphere*, Cambridge University Press, New York, 1961.
- Gherm, V. E., N. Y. Zaalov, N. N. Zernov, and A. V. Nikitin, Statistical characteristics of HF field for the vertical sounding of the ionosphere (in Russian), *Sov. Radiotekh. Elektron.*, 35, 2495-2501, 1990.
- Kravtsov, Y. A., Two new asymptotic methods in a theory of wave propagation in inhomogeneous media (review), *Sov. Phys. Acoust.*, Engl. Transl., 14(1), 1-17, 1968.
- Ludwig, D., Uniform asymptotic expansions at a caustic, *Commun. Pure Appl. Math.*, 19, 215-250, 1966.
- Maslov, V. P., *Perturbation Theory and Asymptotic Methods* (in Russian), Moskov. Gos. Univ., Moscow, 1965. (French translation, J. Lascoux and R. Seneor, Dunod, Paris, 1972.)
- Orlov, Y. I., Asymptotic method for wave field description in arbitrary fluently inhomogeneous media (in Russian), *Sov. Trans. Moskov. Energ. Inst.*, 119, 82-113, 1972.
- Tatarskii, V. I., *Wave Propagation in a Turbulent Medium*, McGraw-Hill, New-York, 1961.
- Tinin, M. V., Wave propagation in a media with large scaled random inhomogeneities (in Russian), *Sov. Radiofizika*, 26, 36-43, 1983.
- Wagen, J.-F., and K. C. Yeh, A numerical study of waves reflected from a turbulent ionosphere, *Radio Sci.*, 21, 583-604, 1986.
- Wagen, J.-F., and K. C. Yeh, Simulation of HF propagation and angle of arrival in a turbulent ionosphere, *Radio Sci.*, 24, 196-208, 1989.
- Yeh, K. C., and C. H. Liu, Radio wave scintillation in the ionosphere, *Proc. IEEE*, 70(4), 324-360, 1982.
- Zernov, N. N., HF wave scattering for the oblique propagation in the ionosphere (in Russian), *Sov. Radiofiz.*, 23, 151-158, 1980.
- Zernov, N. N., Near caustic field in the presence of local inhomogeneities producing significant diffraction effects (in Russian), *Trans. XIV Sov. Conf. Radio Wave Propag., Nauka, Leningrad*, 2, 244-246, 1984.
- Zernov, N. N., The generalization of Rytov's method for the case of a point source field in the inhomogeneous media (in Russian), *Sov. Radiotekh. Elektron.*, 35, 1590-1595, 1990.
- Ziolkowski, R. W., and G. A. Deschamps, Asymptotic evaluation of high-frequency fields near a caustic: An introduction to Maslov's method, *Radio Sci.*, 19, 1001-1025, 1984.
- V. E. Gherm, A. V. Nikitin, N. Yu. Zaalov, and N. N. Zernov, Institute of Physics, University of S. Petersburg, Ulyanovskaya 1, Petrodvorets, 198904, Russia.

On some problems of the theory of radio wave propagation in a randomly inhomogeneous ionosphere

M. V. Tinin, N. T. Afanasyev, S. M. Mikheev, A. P. Pobedina, and O. V. Fridman

Research Institute for Applied Physics, Irkutsk State University, Irkutsk, Russia

(Received February 15, 1991; revised October 7, 1991; accepted October 28, 1991.)

In considering HF propagation in a random inhomogeneous ionosphere it is necessary to take into account regular and random caustics. Regular caustics connected with regular refraction of radio waves in the ionosphere form a skip zone and determine the maximum usable frequency (MUF) and the maximum of the oblique incidence backscatter sounding (OBS) signal. Random fluctuations of ionospheric radio rays "wash out" field enhancement in the vicinity of MUF and the maximum OBS signal. The presence of random caustics results in strong intensity fluctuations of ionospheric radio waves. The above mentioned problems are considered by the interference integral method. A scintillation index formula for strong fluctuations of oblique ionospheric radio waves is obtained. Average intensity and average pulse signal which form in the vicinity of MUF are investigated. The peculiarities of oblique multihop radio wave propagation, taking into account random ionospheric inhomogeneities, terrestrial surface roughness, and caustics focusing in skip distance are discussed.

1. INTRODUCTION

Existing analysis methods of radio wave propagation in the ionosphere usually neglect random ionospheric irregularities [Davies, 1969], with the exception of radio physical investigations of the irregular ionospheric structure (see, for example, Gershman *et al.* [1984] and Yeh *et al.* [1975]) and some dedicated studies [Wagen and Yeh, 1989; Golynsky and Gusev, 1976; Gurevich and Tsedilina, 1985; Kiang and Liu, 1985; Zernov, 1980] where the analysis of the signal behavior uses the methods of phase screen, geometrical optics, or some other methods typical for the upper radio frequency range. At the same time, the ionospheric propagation of HF and adjacent lower edge of VHF radio waves have a number of interesting features which would force researchers to either modify substantially the existing methods or seek new techniques for investigating the signal propagation through a randomly inhomogeneous medium.

Such peculiarities may include the following: First, ionospheric radio waves often undergo strong fluctuations caused by the appearance of random micromultipaths and random caustics.

Second, there exist regular caustics, the skip zone and the maximum usable frequency (MUF)

associated with regular refraction of ionospheric radio waves. The presence of random ionospheric irregularities gives rise to the possible propagation of radio waves at frequencies above MUF.

Last, it is necessary to take into account the scattering from the terrestrial rough surface in the vicinity of the caustic in the case of oblique multihop propagation and of oblique incidence backscatter sounding.

On the basis of the interference integral method this paper develops an approach, making it possible to take into account the above peculiarities of radio wave propagation in a randomly inhomogeneous ionosphere.

2. INTERFERENCE INTEGRAL METHOD

In the interference integral method, which was first suggested by Yu. I. Orlov [Orlov, 1972, 1974; Avdeev *et al.*, 1988], (analogical integral representations were obtained in a different way by Arnold [1982, 1986] and Felsen [1986]) for describing fields in a deterministic smoothly inhomogeneous medium, the solution $U(\mathbf{r})$ of wave equation

$$\Delta U(\mathbf{r}) + k^2 \epsilon(\mathbf{r}) U(\mathbf{r}) = 0 \quad (1)$$

is represented as superposition of partial waves U_s , in terms of which the field of incident wave

$$U(\mathbf{r}) = \int_{-\infty}^{\infty} U_s(\mathbf{r}, s) d^2s \quad (2)$$

Copyright 1992 by the American Geophysical Union.

Paper number 91RS02921.
0048-6604/92/91RS-02921\$08.00

is expanded, where $\epsilon(r)$ is relative dielectric permittivity, $k = \omega/c$ is wave number in free space, $\omega = 2\pi f$ is the frequency of operation, and s is a two-dimensional parameter-vector, in terms of which the expansion is carried out. In particular, if an incident spherical wave is expanded into plane waves, s defines the propagation direction of these waves in the neighborhood of the source.

Substituting (2) into wave equation (1), we obtain the same equation for partial waves U_S . For the ray-type interference integral [Avdeev et al., 1988], by solving this equation using the geometrical optics method, we find

$$U_S(r, s) = A_S(r, s)e^{ik\Phi_S(r, s)} \quad (3)$$

where $\Phi_S(r, s)$ and $A_S(r, s)$ are solutions of the eikonal and transport equations

$$(\nabla\Phi_S)^2 = \epsilon(r) \quad (4)$$

$$2\nabla A_S \nabla\Phi_S + A_S \Delta\Phi_S = 0 \quad (5)$$

As is known [Kravtsov and Orlov, 1980], using a family of ray characteristics of (4) it is easy to find solutions of (4)–(5) and hence determine the interference integral (2)–(3). For the field of a spherical wave propagating in an inhomogeneous ionosphere we have

$$\Phi_S = \int_{l_0}^l \sqrt{\epsilon} dl + s \cdot \rho_0 \quad (6)$$

$$\begin{aligned} A_S &= \frac{gik}{8\pi} Q(s) \frac{\sqrt{|\nabla_x \nabla_y|}}{\sqrt{1-s^2}} \\ &= \frac{gikQ(s)}{8\pi^2 \sqrt{1-s^2}} \sqrt{1 + \left(\frac{d\tilde{y}}{dx}\right)^2 + \left(\frac{d\tilde{z}}{dx}\right)^2} \\ &\cdot \left[\left| \frac{\partial \tilde{z}}{\partial x_0} \frac{\partial \tilde{y}}{\partial y_0} - \frac{\partial \tilde{z}}{\partial y_0} \frac{\partial \tilde{y}}{\partial x_0} \right| \right]^{-1/2} \end{aligned} \quad (7)$$

Here $\rho_0 = (x_0, y_0)$, $Q(s)$ is the directivity pattern, and g takes into account the emission power of the source located in free space at the origin coordinates.

Integration in (6) proceeds along the trajectory $\tilde{z}(x, \rho_0)$, $\tilde{y}(x, \rho_0)$, originating from the point $\tilde{z} = 0$, $\rho = \rho_0$ with direction cosines $s_x, s_y, \sqrt{1-s^2}$ and

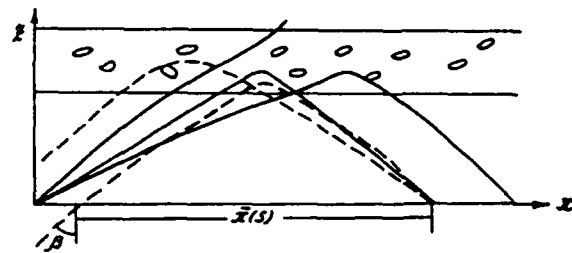


Fig. 1. Geometrical optic rays (solid lines) and ray trajectories of partial waves (dashed lines).

arriving at the observation point $\tilde{x} = x$, $\tilde{z} = z$, $\tilde{y} = y$, from which it follows that

$$x_0(s) = x - \tilde{x}(s), \quad y_0(s) = y - \tilde{y}(s)$$

where $\tilde{x} = \tilde{x}(s)$ and $\tilde{y} = \tilde{y}(s)$ are the x th and y th projections of the hop (projections of ray trajectories of partial waves on the xz plane are shown as dashed lines in Figure 1).

Thus expression (2), in view of (3), (6), and (7), gives an integral representation for the field in a smoothly inhomogeneous medium. By calculating (2) asymptotically by the stationary phase method, one obtains the geometrical optical approximation which becomes inapplicable in the vicinity of the caustics. However, in the vicinity of the caustics of a spherical wave the integral representation (2) can be used if the observer is far away from the reflection point and other caustics of partial plane waves in the vicinity of the observation point are absent because fluctuations of dielectric permittivity are not very large. Using the interference integral it then becomes possible to generalize, in an arbitrary smoothly inhomogeneous medium, the known [Brekovskich, 1970; Felsen and Marcuvitz, 1973] integral representations for the field in a plane-layered medium.

It should be noted that integral representations governing the fields in the vicinity of caustics of different kinds can be obtained using the Maslov method [Kravtsov, 1968]. However, the transition into a mixed coordinate-impulse space assumed in this method, complicates the use of the perturbation theory when determining eikonal fluctuations needed in statistical problems.

The interference integral method, as has become apparent, uses trajectories in the usual coordinate space, and it is possible, when solving (4)–(5), to apply perturbation theory by assuming the fluctua-

tions of dielectric permittivity $\tilde{\epsilon}$ to be small as compared with mean dielectric permittivity $\bar{\epsilon}$: $\epsilon = \bar{\epsilon} + \tilde{\epsilon}$. In this case the interference integral (2) takes on the form

$$U(r) = \int_{-\infty}^{\infty} A_S(r, s) \cdot \exp [ik\Phi_S(r, s) + ik\Phi_S(r, s)] d^2s \quad (8)$$

where $A_S = \bar{A}_S$ and Φ_S is the zero-order (unperturbed) approximation for the amplitude and eikonal of the partial wave obtainable from (6) to (7) when $\epsilon \equiv \bar{\epsilon}$; and Φ_S is the first approximation for the eikonal fluctuations

$$\Phi_S = \frac{1}{2} \int_{l_0}^l \frac{\tilde{\epsilon}}{\bar{\epsilon}} dl \quad (9)$$

Integration in (9) is performed along the unperturbed ray. For amplitude A_S of the partial wave in (8), the zero-order approximation is used because the amplitude fluctuations of the partial waves have a much less influence upon the resulting field as compared with the phase fluctuation.

An asymptotic investigation of (8) has shown [Tinín, 1983] that such an application of perturbation theory in the interference integral method makes it possible to take into account the appearance of random micromultipathing. This property in expression (8), the above noted correctness of describing, using the interference integral method, the effects in the vicinity of the caustics (regular and random) have determined its utilization for taking into account of the peculiarities of radio wave propagation in the randomly inhomogeneous ionosphere, which are given in section 1.

3. IONOSPHERIC RADIO SIGNAL IN THE VICINITY OF MUF

In view of the fact that in the integrand (8) the influence of the irregularities is described by a rather simple (exponential) dependence, using (8) it is easy to find integral representations for the field moments (mean field, mean intensity) and for the intensity dispersion [Avdeev *et al.*, 1988; Tinín, 1983]. Thus, assuming that the path traversed by a partial wave in a randomly inhomogeneous medium, is sufficiently large for normalizing its phase, for the mean intensity we obtain

$$\langle I \rangle = \langle |U|^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A_S(s_1) A_S^*(s_2) \cdot \exp \left\{ ik[\Phi_S(s_1) - \Phi_S(s_2)] - \frac{k^2}{2} D_S(s_1, s_2) \right\} d^2s_1 d^2s_2 \quad (10)$$

where

$$D_S(s_1, s_2) = \langle [\Phi_S(s_1) - \Phi_S(s_2)]^2 \rangle \quad (11)$$

is the structural eikonal function of the partial wave and ensemble averaging denoted by the angle brackets.

By calculating the integral (10) using the stationary phase method we obtain the geometrical optics approximation for the mean intensity tending to infinity as it approaches, in frequency, the MUF (which corresponds in the space to the approach to the caustic).

In the following, for the sake of simplicity, we shall limit ourselves to presenting results for a two-dimensional case, i.e., when transverse changes on the path are negligibly small. In this case, fourfold integral (10) is reduced to a twofold integral, in which variable s_1, s_2 are changed by variable $s = (s_1 + s_2)/2$, $\Delta s = s_1 - s_2$. Then using the asymptotic method [Copson, 1966] the integral along Δs is calculated, and we obtain a one-fold integral:

$$\langle I \rangle = \frac{8\pi^2}{2^{1/3}\kappa^2} \int_{-\infty}^{\infty} \frac{|A_S(s)|^2 s}{\bar{x}(s)\Lambda(s)} \exp \left\{ \frac{4}{3} \left(\frac{\sigma_r(s)}{\Lambda(s)} \right)^6 - \frac{2\sigma_r^2(s)[x - \bar{x}(s)]}{\Lambda^3(s)} \right\} Ai \left\{ \frac{2^{2/3}}{\Lambda(s)} \left[\bar{x}(s) - x + \frac{\sigma_r^4(s)}{\Lambda^3(s)} \right] \right\} ds, \quad (12)$$

where $s = s_x$, $\bar{x}(s)$ is the range of the "hop" of the unperturbed partial ray (see Figure 1) which in the vicinity of the source has the angle $\beta = \arcsin(s)$; $\sigma_r^2(s) = \langle [\partial\Phi_S/\partial s]^2 \rangle$ is the dispersion of this range; $Ai(s)$ is the Airy function;

$$\Lambda(s) = \left(\frac{1}{2k^2} \frac{\partial^2 \bar{x}(s)}{\partial s^2} \right)^{1/3}$$

Expression (12) yields a uniform (on x) asymptotic representation for the mean intensity, valid in the vicinity of the MUF where

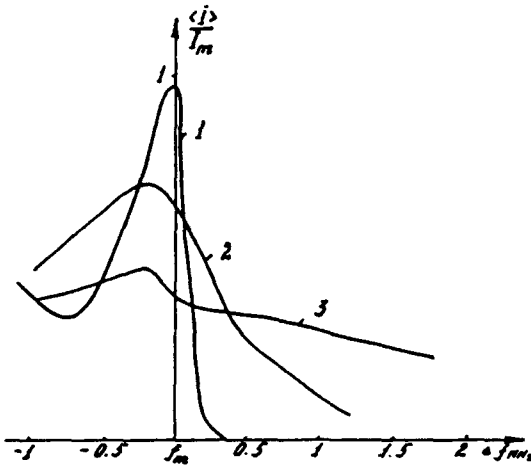


Fig. 2. The behavior of the relative mean intensity $\langle I \rangle / I_m$ of the wave reflected from the ionospheric layer in the vicinity of MUF for $\nu^2 = 10^{-6}$, $l_e = 1$ km (curve 1), $\nu^2 = 10^{-4}$, $l_e = 10$ km (curve 2), $\nu^2 = 10^{-2}$, $l_e = 10$ km (curve 3). I_m is intensity on MUF when random irregularities are absent.

$$x - x_m = \tilde{x}(s_m), \quad \partial \tilde{x}(s_m) / \partial s = 0 \quad (13)$$

If in the region $\tilde{x}(s) \sim x$ that is substantial for integration in (12), the inequality

$$\sigma_x(s) \ll \Lambda(s) \quad (14)$$

is satisfied, which corresponds to the small fluctuations of partial rays within the regular caustic zone, then (12) agrees with the known description of the field in the vicinity of the caustic in terms of the Airy function. If the inverse inequality

$$\sigma_x(s) \gg \Lambda(s) \quad (15)$$

is satisfied, i.e., if the fluctuations of the hop range exceed the width of the caustic zone, then (12) becomes

$$\langle I \rangle = \frac{(2\pi)^{3/2}}{\kappa^2} \int_{-\infty}^{\infty} \frac{|A_s(s)|^2 s}{\sigma_x(s) \tilde{x}(s)} \exp \left\{ -\frac{[x - \tilde{x}(s)]^2}{2\sigma_x^2(s)} \right\} ds \quad (16)$$

Formula (16) describes the summation of intensities of random independent ray fields at the observation point. This physical sense (16) makes it possible to obtain similar expressions in a different way [Lagutkin, 1986]. For illustrating formulas (12) and (16), Figure 2 gives some results of calculations of the mean intensity of the wave reflected from the layer, for which

$$\tilde{\epsilon}(z) = 1 - (f_{cr}/f)^2 \exp \{-[z - z_m]/y_m\}^2 \quad (17)$$

with $y_m = 100$ km, $z_m = 300$ km, and $f_{cr} = 7$ MHz.

On Figure 2 the relative intensity $\langle I \rangle / I_m$ is plotted, where I_m is the intensity of the source on MUF calculated, neglecting random ionospheric irregularities. Quasi-homogeneous random fluctuations were specified by the Gaussian correlation function

$$N(\mathbf{r}_1, \mathbf{r}_2) = \langle \tilde{\epsilon}(\mathbf{r}_1) \tilde{\epsilon}(\mathbf{r}_2) \rangle = \nu^2 [1 - \tilde{\epsilon}(z)]^2 \cdot \exp \{-[(\mathbf{r}_1 - \mathbf{r}_2)/l_e]^2\} \quad (18)$$

Curve 1, calculated by formula (12) for $\nu^2 = 10^{-6}$ and $l_e = 1$ km, corresponds to a weak dispersion σ_x , when the quasi-oscillating dependence of the intensity on the frequency in the illuminated zone (below MUF), which is typical in the absence of irregularities, has not yet been completely "washed out." Curves 2 and 3 are calculated by formula (16) for $\nu^2 = 10^{-4}$, $l_e = 10$ km and $\nu^2 = 10^{-2}$, $l_e = 10$ km, respectively, which already corresponds to the condition (15) for excess effects of random fluctuations of rays over the diffraction penetration of the field into the region of caustic shadow. As is evident from Figure 2, this situation is characterized by the absence of intensity oscillations in the illuminated zone, a decrease of intensity on MUF and by a significantly slower decrease in the region of caustic shadow (i.e., above MUF). Such behavior of the mean intensity in the vicinity of MUF is also evident from the local asymptotic representation for $\langle I \rangle$ obtained through asymptotic calculation of (16) in the vicinity of the skip zone boundary $x \sim x_m$:

$$\langle I \rangle \approx \left(\frac{2\pi}{\kappa} \right) \frac{|A_s(s_m)|^2 s_m}{x_m \sqrt{\sigma_x(s_m) \tilde{x}''(s_m)}} D_{-1/2}(\chi) \exp(-\chi^2/4) \quad (19)$$

where $\chi = (x_m - x)/\sigma_x(s_m)$ and $D_{-1/2}(\chi)$ is the parabolic cylinder function [Beteman and Erdelyi, 1974, p. 125].

On both sides of the caustic, in the illuminated zone and in the region of caustic shadow, one can use asymptotic representations of the functions $D_{-1/2}(\chi)$ [Beteman and Erdelyi, 1974] and from (19) obtain

$$\langle I \rangle \approx \left(\frac{2\pi}{\kappa} \right)^2 \frac{|A_s(s_m)|^2 s_m}{x_m \sqrt{\tilde{x}''(s_m) |x - x_m|}} \exp \left[-\frac{(x - x_m)^2}{2\sigma_x^2(s_m)} \right] \quad (20a)$$

$$x < x_m$$

$$\langle I \rangle \approx \left(\frac{2\pi}{\kappa} \right)^2 \frac{|A_1(s_m)|^2 s_m}{x_m \sqrt{\chi'(s_m)} |x - x_m|} \sqrt{2} \quad x > x_m \quad (20b)$$

From (20b) it follows that in the illuminated zone ($x > x_m$) the mean intensity, as one would expect, becomes a double intensity of the geometrical optics wave, and from (20a) it is evident that in the region of caustic shadow ($x < x_m$) there is an exponential decrease of $\langle I(x) \rangle$. It is interesting to note that in the case considered where condition (15) is satisfied, the rate of this decrease, unlike the case of (14), evidently does not depend on frequency but is determined by the ray dispersion. In view of the frequency dependence of the skip zone range $x_m(f)$, these results can be transformed into the frequency dependence of the mean intensity in the vicinity of the MUF. The integral representations obtained in a similar way for the spatial and frequency coherency functions are useful for investigating the behavior of statistical characteristics of HF signals on frequencies near and above the MUF [Afanasyev *et al.*, 1983]. In turn, using the frequency coherency function, it is possible to obtain an averaged envelope of the impulse signal in the vicinity of the "skip zone boundary." Figure 3 gives some results of a numerical simulation of the behavior of the impulse for the transition through the "skip zone" boundary. Here and in the following the relative intensity $\langle I \rangle / I_1$ is calculated, where I_1 is the intensity of the source in unit distance in free space.

In this case, use is made of an exponential model of dielectric permittivity of the ionosphere (17). The emitted impulse was assumed to be of Gaussian form, with an initial duration of 20 μ s. Random inhomogeneities were specified by a Gaussian correlation function (18) with the parameters $\nu^2 = 10^{-4}$ and $l_c = 1$ km. The parameters of the regular ionospheric layer were chosen the same as in those in Figure 2.

Inspection of Figure 3 reveals that in observation point placed in illuminated zone ($x > x_m = 2251$ km) two signals arrive. These signals corresponding to upper and lower rays [Davies, 1969] which are broader than initial one because of propagation in randomly inhomogeneous ionosphere. Near the "skip zone" boundary ($x \sim x_m$) these signals are merged, and the combined signal is enhanced slightly. In the region of caustic shadow ($x < x_m$) there is a decrease of the combined signal intensity

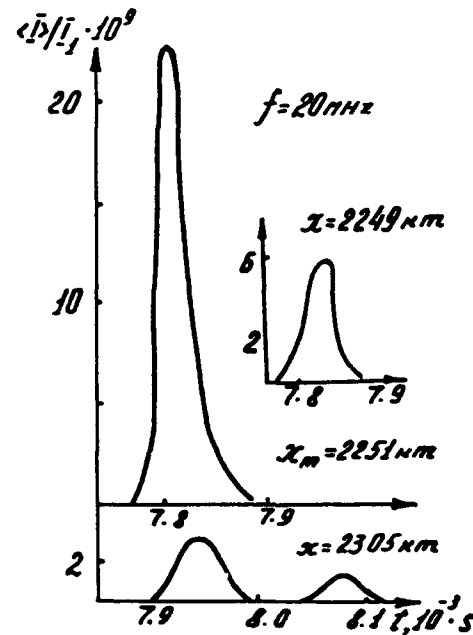


Fig. 3. The behavior of the average impulse for transition through the "skip zone" boundary. Here and in the following I_1 denotes intensity of the source in unit distance.

without substantial additional change of the impulse width.

4. INTENSITY FLUCTUATIONS OF THE WAVE REFLECTED FROM A RANDOMLY INHOMOGENEOUS IONOSPHERE

The property of the interference integral, noted in section 2, correctly allowing for effects of random multipathing, permits us to use this integral for describing strong intensity fluctuations. Further, unlike the known methods of calculating strong intensity fluctuations [Rytov *et al.*, 1971], by using expression (8), it is possible to take into account the regular refraction. In particular, one can obtain an integral representation for the averaged square of the intensity. Next, using the technique for asymptotic calculation of multidimensional integrals as described by Yakushkin [1971], we obtain the following asymptotic (for a large phase dispersion) expression for the scintillation index in a two-dimensional inhomogeneous medium,

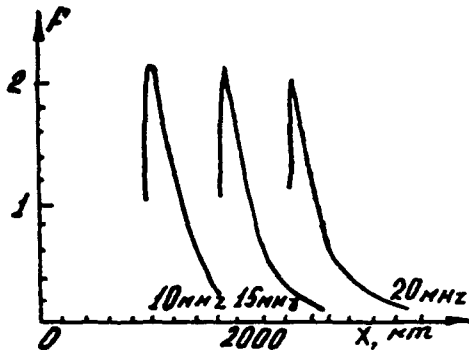


Fig. 4. The scintillation index for low rays.

$$F = \frac{\sigma_I^2}{\langle I \rangle^2} = \frac{\langle I^2 \rangle}{\langle I \rangle^2} - 1$$

$$= \frac{8k}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{|A_s(s)|^4}{\sigma_r(s)} b^2 \exp \left[-\frac{(x - \tilde{x}(s))^2}{2\sigma_r^2} \right] R(s) ds$$

$$\cdot \left\{ \int_{-\infty}^{\infty} \frac{|A_s(s)|^2}{\sigma_r(s)} \exp \left[-\frac{[x - \tilde{x}(s)]^2}{2\sigma_r^2} \right] ds \right\}^{-2} - 1, \quad (21)$$

where

$$R(s) = \frac{\sqrt{\pi}}{4p} \ln(p) e^{-\alpha^2} - \frac{1}{2p} \int_0^x \ln(y) e^{-\alpha^2} \cos(2\alpha y) dy$$

$$- \frac{\sqrt{\pi}}{2p} \int_0^x \ln(x) \frac{d}{dx} [\sqrt{\varphi(x, s)} e^{-\alpha^2} \cos(2\alpha y)] dx + O(p^{-2}) \quad (22)$$

$$\alpha^2 = 0.5[\tilde{x}'(s)]^2/\sigma_r^2(s) \quad (23)$$

$$p^2 = 2k^2\sigma_r^4(s)/\sigma_c^2(s) = k^2\sigma_r^2 b^2 \quad (24)$$

$$b^2 = 2\sigma_r^2(s)/\sigma_c^2 \quad (25)$$

$$\varphi(\eta, s) = \eta^2\sigma_r^2(s)/[\sigma_r^2(s) - \Psi_x(b\eta, s)] \quad (26)$$

$$\sigma_r^2 = \langle \tilde{x}^2(s) \rangle = \langle (\partial\Phi_s/\partial s)^2 \rangle \quad (27)$$

$$\sigma_c^2 = \langle (\partial\tilde{x}/\partial s)^2 \rangle \quad (28)$$

$$\psi_x[s_1 - s_2, (s_1 + s_2)/2] = \langle \tilde{x}(s_1)\tilde{x}(s_2) \rangle \quad (29)$$

is the correlation function of partial ray fluctuations.

In the illuminated region there exist solutions of

the equation for the ray connecting the source and the observation point

$$\tilde{x}(s_i) = x \quad (30)$$

Hence, upon an asymptotic calculation of the integrals in (21), we obtain

$$F = \frac{8}{\pi} \left[\sum_i \alpha(s_i) p(s_i) R(s_i) I_0^2(s_i) \right] \left[\sum_i I_0(s_i) \right]^{-2} - 1, \quad (31)$$

where I_0 is the geometrical optics approximation for the intensity and the summation is carried out over all rays connecting the source and the observation point. Expressions (21) and (31) permit us to investigate the influence of the regular refraction (focusing and defocusing) upon statistical focusing effects and upon saturation of the intensity fluctuations.

In the case of single-ray propagation, instead of (31), we obtain

$$F = \frac{8}{\pi} \alpha(x) p(x) R(x) - 1, \quad (32)$$

where, according to (30), an unambiguous transition is made from the parameter s to the path length x . Considering this case makes sense because under real conditions it is possible to measure F for separate modes. Figures 4 and 5 present the results of the calculations of the scintillation index obtained for a number of operational frequencies f , with background ionosphere parameters determined above in (17), and the random fluctuations, specified by the correlation function (18), with having the scale $l_e = 10\sqrt{2}$ km and intensity $\nu = 0.05$. The curves in Figure 4 correspond to the propagation by the lower ray; Figure 5 presents the Peder-sen ray (upper ray).

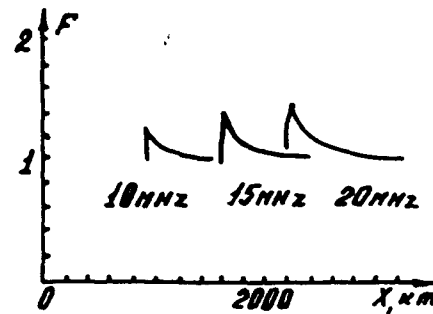


Fig. 5. The scintillation index for upper rays.

Inspection of Figure 4 reveals that in the region of lower rays the scintillation index increases from values close to 0, corresponding to reflection of the rays from the lower ionospheric boundary, at large distances to a certain maximum in the enhancement region. The value of this maximum depends somewhat on the parameters of random irregularities and signal frequencies. This maximum is located near the boundary of the illuminated zone, and the distance from this boundary increases with increasing ν . As the observer shifts to the illuminated zone boundary, F decreases to unity.

The region of upper rays (see Figure 5) also includes a maximum scintillation index near the illuminated zone boundary. Then with an increase in distance the ionospheric path of the upper ray is raised and the scintillation index decreases to unity in the saturation region.

5. INFLUENCE OF RANDOM IONOSPHERIC INHOMOGENEITIES UPON THE STRUCTURE OF THE OBLIQUE INCIDENCE BACKSCATTER SOUNDING (OBS) SIGNAL

Observations of the signal which, after reflection from the ionosphere, is backscattered from the roughness of the terrestrial surface and returns (after a second reflection from the ionosphere) to the observation point, form the basis of the OBS method. When calculating the OBS signal structure, for describing ionospheric radio waves, the geometrical optics approximation is usually used [Croft, 1967]. Backscattering from the vicinity of the "skip zone" gives crucial contribution in the signal [Chernov, 1971]. However, the "skip zone" is known as a region where focusing effects take place, and hence application of the geometrical optics approximation becomes incorrect. There are also publications (see, for example, Dubrovsky *et al.* [1988]) in which, using geometrical optics modifications valid in the vicinity of the caustic, correct allowance for focusing effects is made, but random fluctuations are neglected in this case. Using the interference integral method, it was possible to take into account random ionospheric inhomogeneities and caustic effects in the vicinity of the "skip zone" boundary [Tinín *et al.*, 1988]. In particular, by neglecting correlation of the waves, propagating in the forward and backward directions, and taking into account spherical Earth for an averaged envelope of the narrow-band OBS signal, we obtain the expression

$$\begin{aligned} \langle I(t) \rangle = & \frac{(2\pi)^4}{k^4} \sqrt{\frac{2}{\pi}} \int \int \int_{-\infty}^{\infty} I_0(t') \\ & \cdot \frac{\sigma(s_1, s_2) |A_{s1}(s_1) A_{s2}(s_2)|^2 s_1 s_2}{\sigma_r \Lambda_b R_e^2 \sin [\tilde{x}(s_1)/R_e] \sin [\tilde{x}(s_2)/R_e]} \\ & \cdot \exp \left\{ -0.5[\tau(s_1, s_2) - t + t']^2 \sigma_t^{-2} - 2\gamma(s_1) \right. \\ & \left. - 2\gamma(s_2) + 4\sigma_r^2 \frac{\tilde{x}(s_1) - \tilde{x}(s_2)}{\Lambda_b^3} + \frac{16}{3} \left(\frac{\sigma_x}{\Lambda_b} \right)^6 \right\} \\ & \cdot Ai \left[2 \frac{\tilde{x}(s_1) - \tilde{x}(s_2)}{\Lambda_b} \right. \\ & \left. + 4 \left(\frac{\sigma_x}{\Lambda_b} \right)^4 \right] ds_1 ds_2 dt'. \end{aligned} \quad (33)$$

where $\sigma(s_1, s_2)$ is the scattering cross section of the terrestrial surface roughness averaged over the azimuth, with proper account of the azimuthal dependence of the directivity diagrams of the receiver and transmitter antennas; R_e is the Earth's radius; $\chi(s_1)$ is the absorption of the wave directed to the scattering region; $\chi(s_2)$ is the absorption of the wave returning from the scattering region; $\sigma_x^2(s_1, s_2) = \sigma_{x1}^2(s_1) + \sigma_{x2}^2(s_2)$ is a total dispersion of the fluctuations of the hop range; $\tau(s_1, s_2) = \tau(s_1) + \tau(s_2)$ is a total time of wave propagation to and from the scattering region; $\sigma_t^2(s_1, s_2) = \sigma_{t1}^2(s_1) + \sigma_{t2}^2(s_2)$ is dispersion of this propagation time; $\Lambda_b(s_1, s_2) = |W|^{1/3} \text{sgn}(W)$, $W = [\tilde{x}_1'(s) - \tilde{x}_2'(s)]/k^2$; and $I_0(t')$ is the envelope of the emitted pulse.

As in the case of oblique-incidence sounding described in section 3, with a greater dispersion of rays, namely when

$$\sigma_x > |\Lambda_b| \quad (34)$$

interference effects in the OBS signal structure are "washed out," and (33) takes the form

$$\begin{aligned} \langle I(t) \rangle = & \frac{(2\pi)^3}{k^4} \int \int \int_{-\infty}^{\infty} I_0(t') \\ & \cdot \frac{\sigma(s_1, s_2) |A_{s1}(s_1) A_{s2}(s_2)|^2 s_1 s_2}{\sigma_x \sigma_r R_e^2 \sin [\tilde{x}(s_1)/R_e] \sin [\tilde{x}(s_2)/R_e]} \\ & \cdot \exp \left\{ -2\gamma(s_1) - 2\gamma(s_2) \right. \\ & \left. - \frac{[\tau(s_1, s_2) - t + t']^2 [\tilde{x}(s_1) - \tilde{x}(s_2)]^2}{2\sigma_t^2} \right\} ds_1 ds_2 dt' \end{aligned} \quad (35)$$

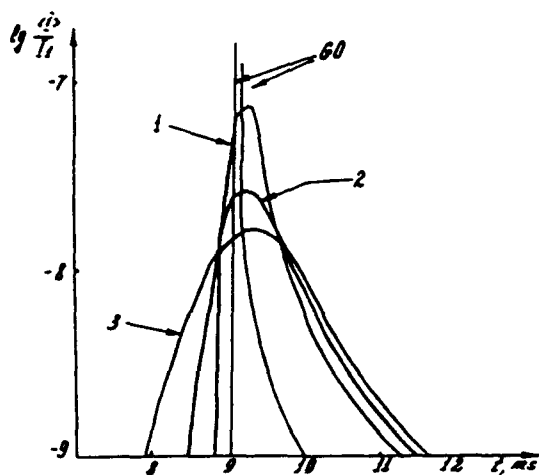


Fig. 6. The average envelope of the oblique incidence-backscatter sounding signal for $\nu^2 = 10^{-4}$ (curve 1), $\nu^2 = 3 \cdot 10^{-4}$ (curve 2), and $\nu = 5 \cdot 10^{-4}$ (curve 3), $l_e = 5$ km. Curve GO is the geometrical optics approximation neglecting random inhomogeneities.

In order to illustrate the OBS theory developed we shall give some results of simulations of the OBS signal structure on a one-hop path in the case of a horizontally homogeneous one-layer ionosphere without allowance for the absorption. As in the case of section 3, the height distribution of dielectric permittivity of the ionized layer $\bar{\epsilon}(r)$ was specified by expression (17), and the correlation function of random fluctuations $N(r_1, r_2)$ was specified by expression (18). The calculations assumed $l_e = 5$ km. The emitted impulse was assumed to be of rectangular form with an initial duration of 0.3 ms.

Figure 6 gives the envelope of the OBS signal for a different degree of ionospheric disturbance: curve 1 corresponds to plasma frequency fluctuations making up 1% of the background disturbance, curve 2, 3%, and curve 3, 5%. This example indicates that the presence of random ionospheric inhomogeneities leads to washing-out of the forward and backward boundaries of the signal by a value of $\sim \sigma_i$ and to a decrease in amplitude. For comparison, the signal calculated in the geometrical optics (GO) approximation is also presented. The signal strength becomes infinity after the moment of arriving of the rays backscattered from the "skip zone." By comparing with GO calculations, which are useful for inferring the arrival time of the forward front, and the amplitude in the backward part of the signal in the absence of inhomogeneities, it becomes apparent

that, for improving the accuracy of calculating a minimum delay time it is necessary to take into account the effect of random inhomogeneities.

6. SOME EFFECTS OF RANDOM IONOSPHERIC INHOMOGENEITIES IN THE CASE OF MULTIHOP PROPAGATION

The roughness of the terrestrial surface should be taken into account, not only in the case of backscattering as done in the preceding section but also in the case of multihop propagation, when the main role is played by direct forward scatter. This can also involve situations when the interference integral method must be used. In order to see this we consider, initially, the propagation process in the geometrical optics approximation when the waves arrive at the observation point after double reflection from the layer and one reflection from the surface (see Figure 7). In this case, using the Kirchhoff approximation [Rytov *et al.*, 1978], when taking into account the direct forward scattering on the rough surface, for the mean intensity we obtain

$$\langle I \rangle = \int I_{01}(\eta) I_{02}(\eta) |Vq_z|^2 \sigma_B^{-1} \exp[-q_z^2 / (2\sigma_B^2)] d\eta. \quad (36)$$

where $I_{01}(\eta)$ and $I_{02}(\eta)$ are the intensities of geometrical optic waves in the absence of fluctuations on the first and second hops, respectively; V is Fresnel's reflection coefficient; $q = (q_x, q_z)$ is the wave vector of reflection from the surface, over which integration is performed in (36); and σ_B^2 is the dispersion of its x th component, with allowance for the scattering due to the layer's irregularities and the surface roughness.

Formula (36) describes the incoherent summation of the intensity of all waves arriving at the observation point after being scattered from the surface roughness; it is also clear that for these waves, in the general case, the angle of incidence on the surface does not equal the reflection angle (see Figure 7).

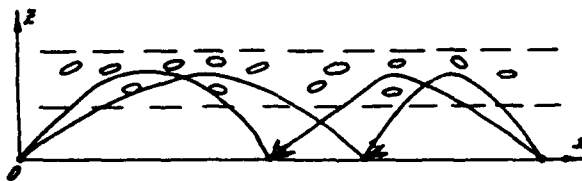


Fig. 7. Two-hop propagation taking into account the direct forward scattering on the rough Earth surface.

The integrand in (36) has a maximum when $q_x = 0$, i.e., for those rays for which the laws of mirror reflection hold true. In view of this, as a result of the asymptotic calculation of the integral (36), we obtain the expression for the mean intensity which coincides with the formula for the intensity without allowance for effects of surface roughness and the layer's random inhomogeneity. However, when these effects are taken into account, there appears to be some scattering cone, within which it is necessary to take into account the behavior of intensities $I_{01}(\eta)$, $I_{02}(\eta)$, which is characterized by the presence of an abrupt increase in the vicinity of the caustics of the rays emitted from the transmitting and receiving sites. The presence of such "bright" regions (regions of field swelling) leads to the fact that with increasing scattering cones, apart from rays satisfying the condition $q_x = 0$, a large contribution to the intensity of rays also is made by two groups of rays, one of which consists of rays arriving in the vicinity of the first caustic, and the second one consists of rays arriving from the vicinity of the second caustic.

Expression (36) is unsuitable for taking these effects into account because in the geometrical optics approximation, in the "bright" regions of our interest the intensity of the geometrical optic waves goes to infinity. It is possible to generalize expression (36) so as to correctly describe the scattering effects on the surface roughness in the vicinity of the caustics, by using the integral representation (2). As a result (for details see the work by Tinin [1989]), we obtain

$$\langle I \rangle = \frac{(2\pi)^2}{k^4} \iint_{-\infty}^{\infty} \frac{|V_s q_z(s_1, s_2) A_{s1}(s_1) A_{s2}(s_2)|^2}{\sqrt{\Delta_x} [\bar{x}(s_1)/s_1 + \bar{x}(s_2)/s_2]} \cdot \exp \left\{ -\frac{1}{2\Delta_x} (q_x^2 \sigma_x^2 + 2q_x x_0 \Psi_{x\beta} + x_0^2 \sigma_\beta^2) \right\} ds_1 ds_2 \quad (37)$$

where $A_{s1}(s_1)$ and $A_{s2}(s_2)$ are the amplitudes of partial plane waves on the first and second hops, respectively; and $\Psi_{x\beta}$ is the correlation function between the fluctuations of the ray range and of scattering angles;

$$\Delta_x = \sigma_\beta^2 \sigma_x^2 - \Psi_{x\beta}^2; \quad x_0 = x - \bar{x}_1(s_1) - \bar{x}_2(s_2)$$

Expression (37) and similar expressions for the coherency function make it possible to correctly

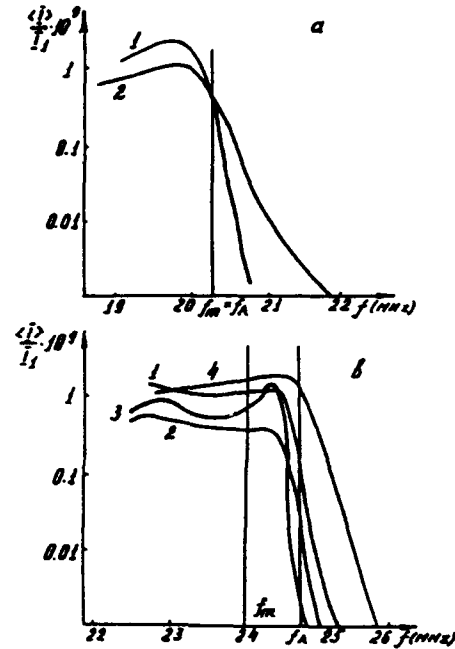


Fig. 8. The dependence of the mean intensity of the wave on frequency on a two-hop path, depending on parameters of ionospheric irregularities (ν_i , l_i) and terrestrial surface roughness (ν_g , l_g): (curve 1) $\nu_i = 10^{-3}$, $l_i = 1$ km, $\nu_g = 100$ m, $l_g = 10$ km; (curve 2) $\nu_i = 10^{-3}$, $l_i = 1$ km, $\nu_g = 100$ m, $l_g = 1$ km; (curve 3) $\nu_i = 10^{-4}$, $l_i = 1$ km, $\nu_g = 100$ m, $l_g = 10$ km; (curve 4) $\nu_i = 10^{-2}$, $l_i = 1$ km, $\nu_g = 100$ m, $l_g = 10$ km; (a) $k_x = 0$, (b) $k_x = 4$ MHz/1000 km.

take into account caustic focusing effects in the case of propagation between a randomly inhomogeneous layer and the surface roughness.

Figure 8 presents the results of a numerical simulation using formula (37), of the intensity of the HF signal field on a two-hop path, depending on parameters of ionospheric irregularities and of terrestrial surface roughness. As the ionospheric layer model, the dependence of average dielectric permittivity on coordinates of the form:

$$\bar{\epsilon}(z, x) = 1 - [(f_{cr} + K_x x)/f]^2 \exp \{ -[(z - z_m)/y_m]^2 \}$$

is used here, where K_x is the critical frequency gradient.

Ionospheric irregularities were defined by Gaussian correlation function of fluctuations (18). Following Chernov [1971], correlation function of terrestrial surface roughness also was approximated by Gaussian function (18). The calculations were carried out for a path of length $D = 5000$ km.

depending on the working frequency for the following parameters of the ionospheric layer: $f_{cr} = 6.7$ MHz, $z_m = 300$ km, and $y_m = 100$ km. Also, Figure 8a corresponds to the case $K_x = 0$, and Figure 8b corresponds to the case $K_x = 4$ MHz/1000 km. As follows from Figure 8a, the frequency variation of the field intensity in the absence of the horizontal gradient qualitatively resembles the field intensity behavior near the MUF on a one-hop radio path (see Figure 2). The horizontal gradient of critical frequency leads to a more complex frequency dependence of the field intensity (see Figure 8b).

In Figure 8 the frequency f_m is two-hop MUF, calculated by the ordinary method without allowance for effects of surface roughness and ionospheric random irregularities. The frequency f_A is two-hop path MUF calculated by the method used by Agarishev [1985]. In this paper it is supposed and demonstrated experimentally that in the case of multihop ionospheric propagation of radio waves, MUF is determined from the condition of MUFs equality on one-hop paths, composing the all path. Evidently, it corresponds to the matching conditions for "bright" regions which are discussed above. From our theoretical results (see Figure 8b, curves 1-3) it is easy to see that f_A more than f_m corresponds to two-hop MUF, when there are horizontally inhomogeneous layers and random ionospheric irregularities and terrestrial surface roughness. However, it should be noted that for intensive ionospheric irregularities radio wave propagation above frequency f_A (see curve 4 on Figure 8b) is possible.

7. CONCLUSIONS

As is apparent from the preceding discussion, the interference integral method is useful for obtaining integral representations for different moments of the field of waves propagating through a randomly inhomogeneous ionosphere. Using these integral representations, it becomes possible to make asymptotic and numerical analyses of the space-time structure of radio waves reflected from the randomly inhomogeneous ionosphere. Even in poorly studied regions, namely in the vicinity of the MUF and above, under conditions of strong intensity fluctuations, during oblique incidence-backscatter ionospheric sounding and distant multihop propagation in a waveguide produced between randomly

inhomogeneous ionosphere and the rough terrestrial surface, these analyses can be made.

In the calculations a number of assumptions have been made, namely of the Gaussian character of the irregularity spectrum, a two-dimensional picture of the propagation, etc. These assumptions are not crucial but are made temporarily in order to simplify the preliminary calculations. There are more serious difficulties, we believe, which arise when applying the interference integral (8) in regions in which perturbation theory is no longer applicable and caustics of partial waves appear. In these regions it seems appropriate to use, instead of the interference integral of the simplest ray type used here, the more complex integrals of Avdeev *et al.* [1988] and Zernov [1990].

Acknowledgments. The reviewers' comments helped improve the presentation of this paper and are very much appreciated.

REFERENCES

- Afanasyev, N. T., V. P. Grozov, A. A. Krasikov, V. E. Nosov, and M. V. Tinin, The function of a frequency correlation of field fluctuations and the mean intensity of the impulsive signal at oblique ionospheric sounding (in Russian), *Issled. Geomagn. Aeron. Fiz. Solntsa*, 63, 180-189, 1983.
- Agarishev, A. I., Method of maximum usable frequency calculation for two-hop HF propagation, *Radiotekhnika*, 4, 67-70, 1985.
- Arnold, J. M., Oscillatory integral theory for uniform representation of wave function, *Radio Sci.*, 17, 1181-1191, 1982.
- Arnold, J. M., Geometrical theories of wave propagation: A contemporary review, *IEE Proc., Part I: Solid State Electron Devices*, 133, 165-188, 1986.
- Avdeev, V. B., A. V. Demin, Yu. A. Kravtsov, M. V. Tinin, and A. P. Yarygin, The method of interference integrals (A review), *Isv. Vyssh. Uchebn. Zaved Radiofiz.*, 31(11), 1279-1294, 1988.
- Beteman, J., and A. Erdelyi, *Higher Transcendental Functions*, Part II, 296 pp., Nauka, Moscow, 1974.
- Brekovskich, L., *Waves in Layered Media*, Academic, San Diego, Calif., 1970.
- Chernov, Yu. A., *Backscatter Probing of the Ionosphere*, 204 pp., Svyaz, Moscow, 1971.
- Copson, E. T., *Asymptotic Expansions*, 160 pp., Mir, Moscow, 1966.
- Croft, J. A., Computation of HF ground backscatter amplitude, *Radio Sci.*, 2, 739-746, 1967.
- Davies, K., *Ionospheric radiowaves*, Blaisdell, Waltham, Mass., 1969.
- Dubrovsky, K. M., E. V. Ipatov, and E. A. Palkin, An investigation of the influence of spatial focusing of HF radio waves upon the structure of amplitude profiles of signals, in *Propagation and Diffraction of Waves*, edited by D. S. Lukin, pp.

- 123-129, Moscow Physical Technical Institute, Moscow, 1988.
- Felsen, L. B., A spectral view of wave propagation, in *URSI Symposium International on Electromagnetic Theory, August 25-29, 1986, Budapest, Hungary*, Part A, pp. 3-5, Akademiai Kiado, Budapest, 1986.
- Felsen, L. B., and N. Marcuvitz, *Radiation and Scattering of Waves*, Prentice-Hall, Englewood Cliffs, N. J., 1973.
- Gershman, B. N., L. M. Yerukhimov, and Yu. Ya. Yashin, *Wave Phenomena in the Ionospheric and Cosmical Plasmas*, Nauka, Moscow, 1984.
- Golynsky, S. M., and V. D. Gusev, Statistics of rays in a medium with ellipsoidal irregularities and with regular refraction, *Geomagn. Aeron., Engl. Transl.*, 16, 1026-1031, 1976.
- Gurevich, A. V., and E. E. Tsedilina, *Long Distance Propagation of HF Radio Waves*, Springer-Verlag, New York, 1985.
- Kiang, Y.-W., and C. H. Liu, Wave scattering by an irregularity slab embedded in a stratified medium: Applications to ionospheric propagation, *Radio Sci.*, 20, 80-94, 1985.
- Kravtsov, Yu. A., Two new asymptotic methods in the theory of wave propagation in inhomogeneous media, *Sov. Phys. Acoust., Engl. Transl.*, 14, 1-17, 1968.
- Kravtsov, Yu. A., and Yu. I. Orlov, *Geometrical Optics of Inhomogeneous Media*, Nauka, Moscow, 1980.
- Lagutkin, V. N., Concerning the calculation of statistical characteristics of backscatter signals in a randomly irregular ionospheric waveguide, *Radiotekh. Electron., Moscow*, 10, 1975-1982, 1986.
- Orlov, Yu. I., An asymptotic method for determining the field in an arbitrary smoothly inhomogeneous medium, *Tr. Mosk. Energ. Isnt.*, 119, 82-91, 1972.
- Orlov, Yu. I., An uniform asymptotic integral representation of fields in smoothly inhomogeneous media, *Izv. Vyssh. Uchebn. Zaved Radiofiz.*, 17(7), 1035-1041, 1974.
- Rytov, S. M., Yu. A. Kravtsov, and V. I. Tatarsky, *Introduction to Statistical Radio Physics*, Part II, 463 pp., Nauka, Moscow, 1978.
- Tinin, M. V., Propagation of wave in medium with large-scale random inhomogeneities, *Izv. Vyssh. Uchebn. Zaved Radiofiz.*, 26, 36-43, 1983.
- Tinin, M. V., Propagation of the HF signal between a randomly inhomogeneous ionosphere and rough terrestrial surface (in Russian), *Issled. Geomagn. Aeron. Fiz. Solntsa*, 88, 145-155, 1989.
- Tinin, M. V., S. M. Mikheev, and O. V. Fridman, Structure of radio signal from backscatter probing of the ionosphere, *Izv. Vyssh. Uchebn. Zaved Radiofiz.*, 31(9), 1027-1035, 1988.
- Wagen, J.-F., and K. C. Yeh, Simulation of HF propagation and angle of arrival in a turbulent ionosphere, *Radio Sci.*, 24, 196-208, 1989.
- Yakushkin, I. G., Intensity fluctuations of the plane wave field behind the chaotic phase screen, *Izv. Vyssh. Uchebn. Zaved Radiofiz.*, 17(9), 1350-1355, 1971.
- Yeh, K. C., C. H. Liu, and M. Y. Youakim, A theoretical study of the ionospheric scintillation behaviour caused by multiple scattering, *Radio Sci.*, 10, 97-106, 1975.
- Zernov, N. N., Scattering of waves of the SW range in oblique propagation in the ionosphere, *Radiophys. Quantum Electron., Engl. Transl.*, 23(2), 151-158, 1980.
- Zernov, N. N., Rytov's method generalization for the case of point source in inhomogeneous media, *Radiotekh. Electron., Moscow*, 35(8), 1590-1595, 1990.
- N. T. Afanasyev, O. V. Fridman, S. M. Mikheev, A. P. Pobedina, and M. V. Tinin, Research Institute for Applied Physics, Irkutsk State University, b. Gagarina, 20, Irkutsk, Russia.

On three-dimensional propagation and dispersion of wave packets described by spatial moments

B. Dong¹ and K. C. Yeh

Wave Propagation Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign

(Received February 16, 1991; revised May 16, 1991; accepted May 23, 1991.)

In this paper we study the three-dimensional propagation of a vector wave packet in a lossless, homogeneous but anisotropic medium. By using the superposition principle and Fourier transform theory a general formulation of wave packet propagation theory is developed. We prove that the square magnitude of the complex amplitude of the wave packet over the whole space integrates to a constant. The conservation of this integral merely reflects the lossless nature of the medium. This integral can be considered as the zeroth spatial moment. Upon this foundation the first and second spatial moments are introduced. The first moment determines the position vector of the centroid of a wave packet. It is a linear function of time, showing the motion of the centroid at a spectrum weighted average of the gradient of frequency in wave number space. Thus the concept of group velocity is extended to the wide banded wave packets. The second moment of the wave packet describes the property of dispersion. It reflects the anisotropic nature of the propagating medium and reveals the fact that the length of the wave packet expands with time in the form of time squared.

1. Introduction

The theory of one-dimensional wave packet propagation has been built on the foundation laid by Sommerfeld and Brillouin [Brillouin, 1960]. The concept of a wave packet has played an important role in various fields ranging from quantum mechanics to atmospheric dynamics. Recently, the wave packet theory has been generalized to a multidimensional vector wave case, and some interesting results have been obtained [Yeh *et al.*, 1991]. In contrast with the monochromatic case the polychromatic property of a wave packet modifies the wave properties in at least three aspects: (1) pulse dispersion, (2) polarization, and (3) energy flow. The dispersion of a pulse in one-dimensional

propagation has been known for a long time [Havelock, 1914]. In three-dimensional propagation in an anisotropic medium, in addition to translation along the group velocity, the pulse dispersion is found to depend on directions. For electromagnetic waves in a magnetoplasma [Xu and Yeh, 1990] and for atmospheric gravity waves [Dong and Yeh, 1989] the direction-dependent stretching on an initial Gaussian pulse can be described as rotation of the major axis of the pulse. In a monochromatic case the locus of the tip of the field vector (e.g., electric field for electromagnetic waves or velocity vector for atmospheric waves) may be described by a closed elliptical curve. On the other hand, the locus of the tip of a polychromatic wave is generally not closed, it is open. Furthermore, the polarization generally depends on the position relative to the center of the pulse and for some positions the sense of rotation can even be reversed. For monochromatic waves the energy may be confined to flow in a certain plane. This is so, for example, for whistlers as their energy flow is confined to the plane defined by the steady magnetic field and the wave number vector. For polychromatic waves there may exist a small energy flow out of the plane. Therefore the three-

¹ On leave from Chengdu Institute of Meteorology, Chengdu, China.

Copyright 1992 by the American Geophysical Union.

Paper number 91RS02112.

0048-6604/91/91RS-02112\$08.00

dimensional packet propagation of vector waves in a dispersive and anisotropic medium can be quite complex.

In this paper a moment approach is developed to deal with the propagation and dispersion problem generally. A moment approach has long been adopted by many authors in the literature. In the field of wave propagation, *Yeh and Liu* [1977] developed the method to investigate pulse distortion in a random medium. They found the use of temporal moments of a signal pulse very convenient in describing the evolution of pulse distortion arising from multiple scattering. Similarly, the moment method has been applied by *Van Groesen and Mainardi* [1989] to a dissipative medium. In quite similar ways the first and second spatial moments of a wave packet are introduced in this work to quantify the wave packet's position, motion, and dispersion. The results show that the moment approach is effective and natural in dealing with wave packet problems. Besides, the results are very general and have clear physical meaning. They are valid in cases of narrow banded as well as wide banded wave packets.

2. Construction of a Wave Packet

We consider an n -dimensional space-time field vector $\tilde{F}(r, t)$ constituted by n physical variables $\tilde{F} = (F_1, \dots, F_n)^T$, which satisfy the following vector first-order differential equation.

$$\frac{\partial}{\partial t} \tilde{F} + [\tilde{H}(\nabla) + \tilde{Q}] \cdot \tilde{F}(r, t) = 0 \quad (1)$$

where \tilde{Q} is a constant $n \times n$ matrix while \tilde{H} is an $n \times n$ first-order spatial differential operator matrix of constant coefficients with $\tilde{H}(0) = \tilde{0}$.

Consider first the monochromatic solution

$$\tilde{F} = e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} \tilde{u} \quad (2)$$

where \tilde{u} is a complex constant vector normalized to satisfy $|\tilde{u}|^2 = \tilde{u}^* \cdot \tilde{u} = 1$. Substituting (2) into (1), we have

$$[j\omega \tilde{I} + \tilde{H}(-j\mathbf{k}) + \tilde{Q}] \cdot \tilde{u} = 0 \quad (3)$$

This is an eigenvalue and eigenvector problem. For nontrivial solutions we equate the determinant of the coefficient matrix of (3) to zero, obtaining the dispersion equation

$$\text{Det} [j\omega \tilde{I} + \tilde{H}(-j\mathbf{k}) + \tilde{Q}] = 0 \quad (4)$$

This is a relation between ω and \mathbf{k} . In general there may exist many roots, each one corresponding to an eigenmode. For each eigenmode, ω can be formally expressed in terms of \mathbf{k} in an explicit form

$$\omega = \omega(\mathbf{k})$$

Corresponding to each eigenfrequency, there exists an eigenvector \tilde{u} also determined by \mathbf{k} .

$$\tilde{u} = \tilde{u}(\mathbf{k})$$

According to the superposition principle, a wave packet can be constructed by the following vector integral in \mathbf{k} space.

$$\tilde{F} = \int S(\mathbf{k}) e^{j\omega(\mathbf{k})t} \tilde{u}(\mathbf{k}) e^{-j\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{k} \quad (5)$$

In the case that $S(\mathbf{k})$ is significant only in the neighborhood of a certain carrier wave number \mathbf{k}_0 the integral (5) could be rewritten as

$$\tilde{F} = e^{j(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r})} \int S(\mathbf{k}_0 + \mathbf{k}) e^{j[\omega(\mathbf{k}_0 + \mathbf{k}) - \omega_0]t} \cdot \tilde{u}(\mathbf{k}_0 + \mathbf{k}) e^{-j\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{k} \quad (6)$$

where $\omega_0 = \omega(\mathbf{k}_0)$. In the following, in the interest of avoiding using too many symbols, we will use $S(\mathbf{k})$ to replace $S(\mathbf{k}_0 + \mathbf{k})$, $\tilde{u}(\mathbf{k})$ to replace $\tilde{u}(\mathbf{k}_0 + \mathbf{k})$, and $\omega(\mathbf{k})$ to replace $\Delta\omega \equiv \omega(\mathbf{k}_0 + \mathbf{k}) - \omega_0$. Occasionally, this new convention will be reminded, especially when there is possibility for confusion. In the new convention we note that $\omega(0) = 0$ and (6) can be written as

$$\tilde{F} = e^{j(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r})} \tilde{J} \quad (7)$$

where

$$\tilde{J} = \int S(\mathbf{k}) e^{j\omega(\mathbf{k})t} \tilde{u}(\mathbf{k}) e^{-j\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{k} \quad (8)$$

Inserting (7) into (1), we obtain the equation satisfied by \tilde{J} .

$$\left[\frac{\partial}{\partial t} \tilde{I} + \tilde{H}(\nabla) + \tilde{L}_0 \right] \cdot \tilde{J} = 0 \quad (9)$$

where

$$\tilde{L}_0 = j\omega_0 \tilde{I} + \tilde{H}(-j\mathbf{k}_0) + \tilde{Q} \quad (10)$$

Equation (8) shows that $\tilde{J}(\mathbf{r}, t)$ is the three-dimensional Fourier transform of a triple product consisting of three factors $S(\mathbf{k})$, $e^{j\omega(\mathbf{k})t}$, and $\tilde{u}(\mathbf{k})$. Let us denote the Fourier transform of $S(\mathbf{k}) e^{j\omega(\mathbf{k})t}$ by $A(\mathbf{r}, t)$, the Fourier transform of $e^{j\omega(\mathbf{k})t}$ by $G(\mathbf{r}, t)$, the Fourier transform of $\tilde{u}(\mathbf{k})$ by $\tilde{U}(\mathbf{r})$. According to the convolution property of Fourier transform theory, \tilde{J} can be expressed as

$$\tilde{J} = A(\mathbf{r}, t) * \tilde{U}(\mathbf{r}) \quad (11)$$

where

$$A(\mathbf{r}, t) = A(\mathbf{r}, 0) * G(\mathbf{r}, t) \quad (12)$$

Notice that $A(\mathbf{r}, 0)$ is the Fourier transform of $S(\mathbf{k})$. Owing to the distributive properties of multiple convolution integrals, \tilde{J} can also be written as

$$\tilde{J}(\mathbf{r}, t) = G(\mathbf{r}, t) * \tilde{J}(\mathbf{r}, 0) \quad (13)$$

where

$$\tilde{J}(\mathbf{r}, 0) = A(\mathbf{r}, 0) * \tilde{U}(\mathbf{r}) \quad (14)$$

We will call $G(\mathbf{r}, t)$ the propagator. It is a Green's function of the initial value problem in both the scalar case (12) and the vector case (13).

Equation (11) can be written out explicitly in terms of a convolution as

$$\tilde{J}(\mathbf{r}, t) = \frac{1}{(2\pi)^3} \int \tilde{U}(\mathbf{r}') A(\mathbf{r} - \mathbf{r}', t) d^3\mathbf{r}' \quad (15)$$

As such, it implies that the value of \tilde{J} at a position \mathbf{r} is determined not only by the value of A at the same point \mathbf{r} but also all the remote points from \mathbf{r} .

If $S(\mathbf{k}) e^{j\omega(\mathbf{k})t}$ is very sharp in the spectrum domain, its Fourier transform $A(\mathbf{r}, t)$ would be very smooth in the configuration space. As a result, A in (15) will contain diminishing higher order derivatives in the neighborhood of $\mathbf{r}' = 0$. In this case we take Taylor's expansion of A to obtain

$$\tilde{J}(\mathbf{r}, t) = \sum_{n=0}^{\infty} \frac{j^n}{n!} \left. \frac{\partial^n \tilde{u}(\mathbf{k})}{\partial k_{i_1} \dots \partial k_{i_n}} \right|_{\mathbf{k}=\mathbf{0}} \frac{\partial^n}{\partial x_{i_1} \dots \partial x_{i_n}} \cdot A(\mathbf{r}, t) \quad (16)$$

Notice that here $\mathbf{k} = \mathbf{0}$ must be interpreted as at the carrier wave number \mathbf{k}_0 since originally $\tilde{u}(\mathbf{k})$ was used to replace $\tilde{u}(\mathbf{k}_0 + \mathbf{k})$. Equation (16) is mathematically equivalent to (15) if all terms in (16)

are kept. Calculations of higher-order derivatives need the information of $A(\mathbf{r}, t)$ in an ever larger zone than those of lower-order derivatives of $A(\mathbf{r}, t)$. Ultimately, the information of $A(\mathbf{r}, t)$ at every point in space is needed since the order of derivatives tends to infinity. In the case that $A(\mathbf{r}, t)$ is very smooth, however, a truncation of the series of (16) to a certain order would be satisfactory in describing the behavior of the wave packet.

3. Centroid and Spectral Weighted Group Velocity

As shown in the last section, complex amplitude $A(\mathbf{r}, t)$ has its Fourier transform $S(\mathbf{k}) e^{j\omega(\mathbf{k})t}$, or

$$A(\mathbf{r}, t) = \int S(\mathbf{k}) e^{j\omega(\mathbf{k})t} e^{-j\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{k} \quad (17)$$

and therefore by Fourier inversion

$$S(\mathbf{k}) e^{j\omega(\mathbf{k})t} = \frac{1}{(2\pi)^3} \int A(\mathbf{r}, t) e^{j\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{r} \quad (18)$$

At the instant $t = 0$,

$$A(\mathbf{r}, 0) = \int S(\mathbf{k}) e^{-j\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{k} \quad (19a)$$

$$S(\mathbf{k}) = \frac{1}{(2\pi)^3} \int A(\mathbf{r}, 0) e^{j\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{r} \quad (19b)$$

Consider the Fourier transform of $|A(\mathbf{r}, t)|^2$

$$\begin{aligned} E(\mathbf{k}, t) &\equiv \frac{1}{(2\pi)^3} \int |A(\mathbf{r}, t)|^2 e^{j\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{r} \\ &= \int S^*(\mathbf{k}') S(\mathbf{k}' + \mathbf{k}) e^{j[\omega(\mathbf{k}' + \mathbf{k}) - \omega(\mathbf{k}')]t} d^3\mathbf{k}' \end{aligned} \quad (20)$$

In particular at $t = 0$, (20) reduces to

$$E(0, t) = \int |S(\mathbf{k})|^2 d^3\mathbf{k} = E(0, 0) = \text{const} \quad (21)$$

It follows that the integral

$$W = \int |A(\mathbf{r}, t)|^2 d^3\mathbf{r} = (2\pi)^3 E(0, t) \quad (22)$$

conserves. The quantity W is nothing but the total energy of the wave packet except for a factor $1/(2\pi)^3$. In fact, in many cases the square of the magnitude of vector \tilde{F} is viewed as the energy density. Thus its integral over the whole space is the total energy

$$\begin{aligned}
\int \tilde{F}^* \cdot \tilde{F} d^3r &= \int \tilde{J}^* \cdot \tilde{J} d^3r \\
&= \int (A^* \cdot \tilde{U}^*) \cdot (A \cdot \tilde{U}) d^3r \\
&= \int S^*(k) \tilde{U}^*(k) \cdot S(k) \tilde{U}(k) d^3k \\
&= \int |S(k)|^2 d^3k = \frac{1}{(2\pi)^3} W
\end{aligned}$$

For later applications it is convenient to normalize the amplitude so that

$$\int |a(r,t)|^2 d^3r = 1 \quad (23)$$

where the amplitude is normalized by rescaling,

$$a(r,t) = \frac{1}{\sqrt{W}} A(r,t)$$

After normalization, the time-dependent position vector of the centroid of a wave packet can be defined by

$$R(t) = \int r |a(r,t)|^2 d^3r \quad (24)$$

From (20) and the properties of Fourier transform, R can be calculated by differentiating $E(k,t)$ at $k = 0$. Thus

$$\begin{aligned}
R(t) &= \frac{1}{W} \int r |A(r,t)|^2 d^3r \\
&= -j \frac{(2\pi)^3}{W} \nabla_k E(k,t) \Big|_{k=0} \\
&= \frac{(2\pi)^3}{jW} \int S^*(k') [\nabla_k S(k') \\
&\quad + jt S(k') \nabla_k \omega(k')] d^3k' = R_0 + t V_g \quad (25)
\end{aligned}$$

where the initial position of the packet is

$$\begin{aligned}
R_0 &= \frac{(2\pi)^3}{jW} \int S^* \nabla_k S d^3k \\
&= \frac{(2\pi)^3}{2jW} \int \nabla_k |S|^2 d^3k + \frac{(2\pi)^3}{W} \int |S|^2 \nabla_k \phi d^3k \quad (26)
\end{aligned}$$

Here ϕ is the argument of complex initial spectrum S . The first term on the right-hand side of (26) vanishes since it can be converted into a surface integral at infinity which is always assumed to be zero. We

shall denote the weighted average of any function $f(k)$ over the wave packet spectrum in k space by $\langle f \rangle$, namely,

$$\langle f \rangle = \int \frac{(2\pi)^3}{W} |S|^2 f(k) d^3k$$

Then we have

$$R_0 = \langle \nabla_k \phi \rangle$$

$$V_g = \langle \nabla_k \omega \rangle \quad (27)$$

According to (25) the centroid of a wave packet moves with the weighted group velocity V_g . As the wave packet propagates, it of course undergoes dispersion and hence distortion. But V_g is weighted only by the spectrum of the initial packet. Once the initial spectrum $S(k)$ of the packet is known, the spectrum weighted group velocity V_g is uniquely defined, regardless of the later distortion of the packet. In the literature the group velocity of a wave packet is usually taken to be $\nabla_k \omega|_{k=k_0}$, the gradient of frequency over wave number at the carrier wave number. The choice of the carrier wave number k_0 , however, is left unspecified. One might argue that since the spectrum of the wave packet is concentrated within a very narrow region in k space an arbitrary choice of carrier wave number k_0 within the band would not bring about much discrepancy. But how narrow must the spectrum be in order for this justification to hold? This of course depends also on how fast $\omega(k)$ varies within the narrow spectrum. We give a one-dimensional example below to illustrate and amplify this argument.

Consider a one-dimensional Gaussian wave packet at $t = 0$ of the form

$$A(x,0) = \frac{1}{(2\pi\sigma^2)^{1/4}} e^{-x^2/4\sigma^2} \quad (28)$$

where A satisfies the one-dimensional equivalent of the normalization condition (23) since $W = \int |A(x,0)|^2 dx = 1$. The initial spectrum for this Gaussian packet is

$$S(k) = \frac{1}{2\pi} \int A(x,0) e^{jkx} dx = \frac{\sqrt{2\sigma}}{(2\pi)^{3/4}} e^{-\sigma^2 k^2} \quad (29)$$

Since $S(k)$ is real, the initial position of this packet is at the origin, i.e., $R_0 = 0$. Let the original dispersion relation be

$$\omega = \alpha k + \beta k^3 \quad (30)$$

At carrier wave number the carrier frequency is

$$\omega_0 = \alpha k_0 + \beta k_0^3$$

The corresponding group velocity at the carrier wave number is

$$\left. \frac{d\omega}{dk} \right|_{k=k_0} = \alpha + 3\beta k_0^2$$

The frequency deviation from the carrier frequency is

$$\Delta\omega = \omega - \omega_0 = (\alpha + 3\beta k_0^2)k + 3\beta k_0 k^2 + \beta k^3$$

and its derivative is

$$\frac{d\Delta\omega}{dk} = (\alpha + 3\beta k_0^2) + 6\beta k_0 k + 3\beta k^2 \quad (31)$$

The velocity of the centroid is given by the spectrum weighted group velocity (27), where ω must be replaced by $\Delta\omega$. Thus

$$V_g = \int 2\pi |S(k)|^2 \frac{d\Delta\omega}{dk} dk = \left. \frac{d\omega}{dk} \right|_{k=k_0} + \frac{3\beta}{4\sigma^2} \quad (32)$$

In (32) the second term contains σ and β . The value of σ as seen in (28) is a measure of the size of wave packet in configuration space (i.e., pulse length) while its reciprocal as seen in (29) is a measure of the spectral width in k space. For the dispersion relation given by (30) the coefficient β reflects the extent of dispersive effects. Therefore (32) suggests that for a wave packet of wide spectrum (small σ^2) if the frequency dispersion is important in the neighborhood of the carrier wave number (large β), the spectrum weighted group velocity is not exactly equal to the frequency gradient at the carrier wave number. The difference between these two quantities is dependent on the spectral width of the initial packet and dispersiveness of the medium. When the analysis is extended to the three-dimensional case this difference includes not only the magnitude of the group velocity but also its direction. It is possible and in some cases convenient to require that the gradient of frequency at carrier wave number, $\nabla_k \omega|_{k_0}$, be equal to the spectrum weighted group velocity, i.e., the velocity of the centroid of a wave packet. In so doing the arbitrariness in the selection of a carrier wave number is removed. For the one-dimensional

example considered above we can modify the carrier wavenumber to $k_0 + \Delta$ so that

$$\nabla_k \omega|_{k_0+\Delta} = V_g = \left. \frac{d\omega}{dk} \right|_{k_0} + \frac{3\beta}{4\sigma^2}$$

The required change in wave number is found to be

$$\Delta = \sqrt{k_0^2 + \frac{1}{4\sigma^2}} - k_0.$$

In the remaining discussion we will always assume that the carrier wave number k_0 is so chosen as to make $V_g = \nabla_k \omega|_{k_0}$ valid.

4. Deviation Tensor and Dispersion of a Wave Packet

In section 3 we have found the centroid of a wave packet moves with the spectrum weighted group velocity. To describe the dispersion of a wave packet, it is useful to define a second-order tensor moment of a wave packet. To make the notations more laconic, we adopt the convention $\overline{f(r)}$ to denote

$$\overline{f(r)} = \int f(r) |a(r,t)|^2 d^3r$$

The deviation tensor is defined as the second centered moment

$$D_{ij}(t) = \overline{(x_i - X_i)(x_j - X_j)} = \overline{x_i x_j} - X_i X_j \quad (33)$$

where X_i is the i th component of $R(t)$. It can be shown that

$$\begin{aligned} \overline{x_i x_j} &= -\frac{(2\pi)^3}{W} \left. \frac{\partial^2}{\partial k_i \partial k_j} E(k,t) \right|_{k=0} \\ &= \left\langle \frac{\partial \xi}{\partial k_i} \frac{\partial \xi}{\partial k_j} \right\rangle + \left\langle \frac{\partial \phi}{\partial k_i} \frac{\partial \phi}{\partial k_j} \right\rangle \\ &\quad + i \left\langle \frac{\partial \phi}{\partial k_i} \frac{\partial \omega}{\partial k_j} + \frac{\partial \phi}{\partial k_j} \frac{\partial \omega}{\partial k_i} \right\rangle + i^2 \left\langle \frac{\partial \omega}{\partial k_i} \frac{\partial \omega}{\partial k_j} \right\rangle \quad (34) \end{aligned}$$

where $\xi = \ln |S|$. Now the Cartesian position of the centroid is (see equations (25) and (27)):

$$X_i = \left\langle \frac{\partial \phi}{\partial k_i} \right\rangle + t \left\langle \frac{\partial \omega}{\partial k_i} \right\rangle$$

which gives the product

$$\begin{aligned} X_i X_j &= \left\langle \frac{\partial \phi}{\partial k_i} \right\rangle \left\langle \frac{\partial \phi}{\partial k_j} \right\rangle \\ &+ t \left(\left\langle \frac{\partial \phi}{\partial k_i} \right\rangle \left\langle \frac{\partial \omega}{\partial k_j} \right\rangle + \left\langle \frac{\partial \phi}{\partial k_j} \right\rangle \left\langle \frac{\partial \omega}{\partial k_i} \right\rangle \right) \\ &+ t^2 \left\langle \frac{\partial \omega}{\partial k_i} \right\rangle \left\langle \frac{\partial \omega}{\partial k_j} \right\rangle \end{aligned}$$

Substitution of the above equation reduces (33) to

$$\begin{aligned} D_{ij}(t) &= \left\langle \frac{\partial \xi}{\partial k_i} \frac{\partial \xi}{\partial k_j} \right\rangle + \left\langle \frac{\partial \phi'}{\partial k_i} \frac{\partial \phi'}{\partial k_j} \right\rangle \\ &+ t \left\langle \frac{\partial \phi'}{\partial k_i} \frac{\partial \omega'}{\partial k_j} + \frac{\partial \phi'}{\partial k_j} \frac{\partial \omega'}{\partial k_i} \right\rangle + t^2 \left\langle \frac{\partial \omega'}{\partial k_i} \frac{\partial \omega'}{\partial k_j} \right\rangle \quad (35) \end{aligned}$$

where, using the convention of summing over repeated index,

$$\phi' = \phi - \left\langle \frac{\partial \phi}{\partial k_i} \right\rangle k_i \quad (36a)$$

$$\omega' = \omega - \left\langle \frac{\partial \omega}{\partial k_i} \right\rangle k_i \quad (36b)$$

These newly defined quantities satisfy the interesting relations

$$\left\langle \frac{\partial \phi'}{\partial k_i} \right\rangle = 0 \quad (37a)$$

$$\left\langle \frac{\partial \omega'}{\partial k_i} \right\rangle = 0 \quad (37b)$$

The physical meaning of (37) can be clarified as follows. In a moving frame with its origin set on the centroid of the wave packet the centered moment (33) in nonmoving frame becomes the regular moment in the moving frame, namely,

$$\overline{(x_i - X_i)(x_j - X_j)} = \overline{x'_i x'_j} \quad (38)$$

where $x'_i = x_i - X_i$ or

$$\mathbf{r}' = \mathbf{r} - \mathbf{R}(t) = \mathbf{r} - \mathbf{R}_0 - \mathbf{V}_g t \quad (39)$$

According to (18)

$$\begin{aligned} S(\mathbf{k}) e^{j\omega(\mathbf{k})t} &= \frac{1}{(2\pi)^3} \int A(\mathbf{r}, t) e^{j\mathbf{k} \cdot \mathbf{r}} d^3 r \\ &= \frac{1}{(2\pi)^3} \int A'(\mathbf{r}', t) e^{j\mathbf{k} \cdot (\mathbf{r}' + \mathbf{R}_0 + \mathbf{V}_g t)} d^3 r' \\ &= e^{j\mathbf{k} \cdot \mathbf{R}_0} e^{j\mathbf{k} \cdot \mathbf{V}_g t} \frac{1}{(2\pi)^3} \int A'(\mathbf{r}', t) e^{j\mathbf{k} \cdot \mathbf{r}'} d^3 r' \\ &= e^{j\mathbf{k} \cdot \mathbf{R}_0} e^{j\mathbf{k} \cdot \mathbf{V}_g t} S'(\mathbf{k}) e^{j\omega' t} \quad (40) \end{aligned}$$

The primed variables in (40) are calculated in the moving frame. Therefore

$$S'(\mathbf{k}) e^{j\omega'(\mathbf{k})t} = |S| e^{j[\phi - \mathbf{k} \cdot \mathbf{R}_0]} e^{j[\omega(\mathbf{k}) - \mathbf{k} \cdot \mathbf{V}_g]t} \quad (41)$$

This leads to (36).

From (36) we can see that in the moving frame we have

$$\nabla_{\mathbf{k}} \omega' \big|_{\mathbf{k}=0} = 0 \quad (42)$$

Since, as discussed earlier, the carrier wave number is determined by the condition $\nabla_{\mathbf{k}} \omega' \big|_{\mathbf{k}=0} = \langle \nabla_{\mathbf{k}} \omega' \rangle$. If the wave packet's spectrum is very narrow, we can approximate the phase variation over the wave spectrum band linearly. As a result, by proper choice of phase reference we can make

$$\frac{\partial \phi}{\partial k_i} = \left\langle \frac{\partial \phi}{\partial k_i} \right\rangle = 0.$$

On the other hand, even for wide band signals, if the initial packet possesses a certain symmetry so that its spectrum S is real, it will result in $\phi = 0$. In both of these cases, (35) becomes

$$D_{ij}(t) = \left\langle \frac{\partial \xi}{\partial k_i} \frac{\partial \xi}{\partial k_j} \right\rangle + t^2 \left\langle \frac{\partial \omega'}{\partial k_i} \frac{\partial \omega'}{\partial k_j} \right\rangle \quad (43)$$

The second term is positive definite. The physical meaning of this fact can be seen by considering the following moment

$$B(\mathbf{n}) = \overline{(\mathbf{r}' \cdot \mathbf{n})^2} \quad (44)$$

where \mathbf{n} is an arbitrary unit vector and can be expressed by its directional cosines $\mathbf{n} = (n_1, n_2, n_3)$.

$$\begin{aligned}
B(n) &= \overline{(x'_i n_i)^2} = n_i n_j \overline{x'_i x'_j} \\
&= \left\langle \frac{\partial \xi}{\partial k_i} \frac{\partial \xi}{\partial k_j} \right\rangle n_i n_j + t^2 \left\langle \frac{\partial \omega'}{\partial k_i} \frac{\partial \omega'}{\partial k_j} \right\rangle n_i n_j \\
&= \left\langle \left(n_i \frac{\partial \xi}{\partial k_i} \right)^2 \right\rangle + t^2 \left\langle \left(n_i \frac{\partial \omega'}{\partial k_i} \right)^2 \right\rangle \quad (45)
\end{aligned}$$

Thus the two terms on the right-hand side of (45) are both positive definite. The positive definiteness of the second term reveals the fact that in any direction in configuration space the length of the wave packet is always increasing with time as time squared.

Until now we have been dealing with the general case. To have some specific results that might be interesting, some approximations should be introduced. We now consider a narrow banded wave packet case in which $|S(k)|$ is significant only in a narrow region in the vicinity of the carrier wave number. Let $P_{ij}(t) = D_{ij}(t) - D_{ij}(0)$. From (43) we obtain

$$P_{ij}(t) = t^2 \int \left[\frac{(2\pi)^3}{w} |S|^2 \right] \frac{\partial \omega'}{\partial k_i} \frac{\partial \omega'}{\partial k_j} d^3k \quad (46)$$

Since $|S|^2$ is narrow banded we can take the two lowest order terms in the Taylor's expansion of the factor $(\partial \omega / \partial k_i) (\partial \omega / \partial k_j)$ about $k = 0$. As mentioned above in the moving frame

$$\left. \frac{\partial \omega'}{\partial k_i} \right|_{k=0} = 0,$$

the lowest-order term is the second term. Thus we have

$$P_{ij}(t) = t^2 \alpha_{il} \langle k_l k_m \rangle \alpha_{mj} = t^2 \alpha_{il} V_{lm} \alpha_{mj} \quad (47)$$

where

$$\begin{aligned}
\alpha_{ij} &= \left. \frac{\partial^2 \omega'}{\partial k_i \partial k_j} \right|_{k=0} \\
V_{lm} &= \langle k_l k_m \rangle
\end{aligned}$$

Or, more laconically, in matrix form

$$\bar{D}(t) = \bar{D}(0) + t^2 \bar{\alpha} \cdot \bar{V} \cdot \bar{\alpha} \quad (48)$$

For the special case of a Gaussian wave packet, more specific and interesting results can be obtained. In that case we assume

$$p(k) = \frac{(2\pi)^3}{W} |S(k)|^2 = \frac{\sqrt{\det \bar{M}}}{(2\pi)^{3/2}} e^{-1/2 m_{ij} k_i k_j} \quad (49)$$

where m_{ij} is an element of matrix \bar{M} which is real, symmetric and positive definite. The term $\det \bar{M}$ denotes \bar{M} 's determinant. The Fourier transform of (49) is

$$q(r) = \int p(k) e^{-jk \cdot r} d^3k = e^{-1/2 m_{ij}^{-1} x_i x_j} \quad (50)$$

Therefore

$$\begin{aligned}
V_{ij} &= \int p(k) k_i k_j d^3k \\
&= - \left. \frac{\partial^2}{\partial x_i \partial x_j} q(r) \right|_{r=0} = m_{ij}^{-1}
\end{aligned}$$

$$D_{ij}(0) = \frac{(2\pi)^3}{W} \int \frac{\partial S}{\partial k_i} \frac{\partial S}{\partial k_j} d^3k = \frac{1}{4} m_{ij}$$

Finally,

$$\bar{D}(t) = \frac{1}{4} \bar{M} + t^2 \bar{\alpha} \cdot \bar{M}^{-1} \cdot \bar{\alpha} \quad (51)$$

All the matrices \bar{D} , $\bar{\alpha}$, and \bar{V} are real, symmetric, and positive definite. According to the theorems in algebra, they have real positive eigenvalues and their eigenvectors are orthogonal to each other. At initial time the major orientations of the wave packet are determined by the eigenvectors of \bar{M} . As time goes on the second term in (51) gradually dominates and then the major orientations of the packet are determined mainly by the eigenvectors of matrix $\bar{\alpha} \cdot \bar{V} \cdot \bar{\alpha}$. Since the absolute value of the determinant of a matrix is equal to the product of the absolute value of all its eigenvalues, we can define a "dispersion time" of the wave packet at which the dispersion term $t^2 \bar{\alpha} \cdot \bar{M}^{-1} \cdot \bar{\alpha}$ is of the same order of magnitude as the first term $\bar{M}/4$ in (51). We denote the geometric average of all three eigenvalues of matrix \bar{M} and $\bar{\alpha} \cdot \bar{M} \cdot \bar{\alpha}$ as

$$g_m = \sqrt[3]{\sigma_1 \sigma_2 \sigma_3} = |\det \bar{M}|^{1/3} \quad (52)$$

$$g_d = \sqrt[3]{v_1 v_2 v_3}$$

$$= |\det (\bar{\alpha} \cdot \bar{M}^{-1} \cdot \bar{\alpha})|^{1/3}$$

$$= |\det \bar{\alpha}|^{2/3} / |\det \bar{M}|^{1/3} \quad (53)$$

where $\sigma_1, \sigma_2, \sigma_3$ are the eigenvalues of \bar{M} while v_1, v_2, v_3 are eigenvalues of $\bar{\alpha} \cdot \bar{M}^{-1} \cdot \bar{\alpha}$. On the basis of these discussions the following equation defines t_d .

$$\frac{1}{4} g_m = t_d^2 g_d \quad (54)$$

Therefore

$$t_d = \frac{1}{2} \left| \frac{\det \bar{M}}{\det \bar{\alpha}} \right|^{1/3} \quad (55)$$

In a time of t_d the centroid of the wave packet has moved over a distance $|V_g| t_d$. This distance may be called the dispersion distance d . Thus

$$d = |V_g| t_d = \frac{|V_g|}{2} \left| \frac{\det \bar{M}}{\det \bar{\alpha}} \right|^{1/3} \quad (56)$$

As an example, consider the propagation of acoustic-gravity waves in the atmosphere. A three-dimensional Gaussian gravity wave packet has [see Dong and Yeh, 1989]:

$$\bar{M} = \begin{pmatrix} l_k^2 & 0 & 0 \\ 0 & l_\theta^2 & 0 \\ 0 & 0 & l_\phi^2 \end{pmatrix}$$

$$\bar{\alpha} = \begin{pmatrix} 0 & -\frac{\omega_b}{k_0^2} \cos \theta & 0 \\ -\frac{\omega_b}{k_0^2} \cos \theta & -\frac{\omega_b}{k_0^2} \sin \theta & 0 \\ 0 & 0 & \frac{\omega_b}{k_0^2} \cos \theta \cot \theta \end{pmatrix}$$

Its dispersion time and dispersion distance have been found as follows

$$t_d = \pi T_b \left[\frac{\sin \theta}{\cos^4 \theta} \left(\frac{l_k}{\lambda} \right)^2 \left(\frac{l_\theta}{\lambda} \right)^2 \left(\frac{l_\phi}{\lambda} \right)^2 \right]^{1/3} \quad (57)$$

$$d = \pi \lambda \left[|\tan \theta| \left(\frac{l_k}{\lambda} \right)^2 \left(\frac{l_\theta}{\lambda} \right)^2 \left(\frac{l_\phi}{\lambda} \right)^2 \right]^{1/3} \quad (58)$$

where ω_b and T_b are Brunt-Väisälä circular frequency and period, respectively, k_0 and λ are the wave number and wave length of the carrier wave, respectively, and θ is the zenith angle of k_0 . From (58) and (57) we can see that the closer the wave is to a monochromatic wave (the larger the values of $l_k/\lambda, l_\theta/\lambda, l_\phi/\lambda$) and the nearer is its carrier wave to horizontal propagation (θ close to $\pi/2$), the longer time or distance it will take for the dispersive distortion to emerge.

5. Conclusion

The three-dimensional propagation of a vector wave packet in a homogeneous but anisotropic medium is found conveniently described by spatial moments using the normalized amplitude squared as the weighting function. The zeroth moment is a constant, reflecting the lossless nature of the medium. The first moment gives the time-dependent position of the centroid of the packet. This centroid position moves linearly with time with a spectrum weighted group velocity. The dispersive distortion of the packet is described by a centered second tensorial moment. In a coordinate frame moving with the packet centroid the pulse length is found to stretch with time as time squared. The rate of stretch depends on the direction which gives rise to the appearance of a rotation of the packet envelope. In this connection it is found convenient to define a dispersion time (or a corresponding dispersion distance) as the time (or distance) of propagation within which not much dispersion is expected.

To have a numerical idea about the order of magnitude involved, we apply the results to the propagation of acoustic-gravity waves. Let the Gaussian pulse length be 10 wavelengths long (i.e., $l_k = l_\theta = l_\phi = 10 \lambda$). A gravity wave propagating with a polar angle $\theta = 5^\circ$ has a dispersion time of 140 times the Brunt-Väisälä period (i.e., $t_d/T_b = 1.4 \times 10^2$) and a dispersion distance of 139.4 times the wavelength length (i.e., $d/\lambda = 1.394 \times 10^2$). If, on the other hand, the polar angle of the propagating gravity wave is changed to $\theta = 80^\circ$, the corresponding normalized dispersion time and normalized dispersion distance are changed to $t_d/T_b = 3.226 \times 10^3$ and $d/\lambda = 5.6 \times 10^2$, which are now increased to larger values.

Acknowledgment. This research was supported by the Atmospheric Science Division of the National Science Foundation under grants ATM 87-04947 and ATM 90-16082.

References

- Brillouin, L., *Wave Propagation and Group Velocity*, Academic, San Diego, Calif., 1960.
- Dong, B., and K. C. Yeh, Polarization and dispersion of a Gaussian gravity-wave packet, *Ann. Geophys.*, 7, 53-60, 1989.
- Havelock, T. H., *The Propagation of Waves in Dispersive Media*, Cambridge University Press, New York, 1914.
- Van Groesen, E., and F. Mainardi, Energy propagation in dissipative systems, Part 1, Centrovlocity for linear systems, *Wave Motion*, 11, 201-209, 1989.
- Xu, J. S., and K. C. Yeh, Propagation of a VLH electromagnetic wave packet in a magnetoplasma, *J. Geophys. Res.*, 95(A7), 10,481-10,493, 1990.
- Yeh, K. C., and C. H. Liu, An investigation of temporal moments of stochastic waves, *Radio Sci.*, 12, 671-680, 1977.
- Yeh, K. C., J. S. Xu, and B. Dong., Propagation of a wave packet in a dispersive and anisotropic medium, *J. Electromagn.Waves Appl.*, 5, 221-240, 1991.
-
- B. Dong and K. C. Yeh, Wave Propagation Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

High-latitude ionospheric phenomena diagnostics by high-frequency radio wave propagation observations

D. V. Blagoveshchenskij

Leningrad Institute of Aviation Instrument Making, Leningrad, USSR

L. V. Egorova and V. M. Lukashkin

Arctic and Antarctic Scientific Research Institute, Leningrad, USSR

(Received January 29, 1991; revised September 3, 1991; accepted September 17, 1991.)

With the help of a system of radio paths in the decameter wave band the problem of diagnostics of high-latitude ionospheric phenomena is solved experimentally. The latter are understood to be auroral substorms, main ionization trough, and daytime polar cusp. On the term "cusp" we suggest cusp/cleft region. As a source of information on a phenomenon, the use is made of either the statistical parameters of radio signals at radio channel output or the radio wave propagation parameter by the data of the oblique sounding of ionosphere. The first approach is taken in diagnostics of auroral substorms and main ionization trough. The oblique sounding is used in daytime polar cusp diagnostics. The experimental observations were carried out by two systems of paths: one is situated north of Eastern Siberia and the other encompasses the area north of the European part of the USSR. The range of corrected geomagnetic latitudes covered is $\Phi' = 55^\circ\text{--}75^\circ$. The main conclusion of the paper is that this system of radio paths allows us to conduct qualitative short-time prediction of the said ionospheric phenomena.

1. INTRODUCTION

The ionospheric perturbations, because of the dynamics of the Earth's magnetosphere which in its turn is affected by the physical phenomena on the Sun, manifest themselves first of all at high latitudes [Akasofu, 1968; Eather, 1985; Galperin *et al.*, 1990; Hunsucker, 1991]. The variations of the higher atmosphere are transported from the high latitudes to the low ones and have, as a rule, an adverse effect on the performance of many kinds of radio technical systems, such as radio communication, radio location, direction finding, navigation, etc. [Hunsucker, 1991]. To reduce these negative effects, one should have reliable data on the pending anomalous events. Monitoring the near-Earth medium provides continuous current information on its state. Therefore the diagnostics of the medium and the prediction based on such data are an important means of increasing the reliability of radio systems operation [Blagoveshchenskij and Zherebtsov, 1987; Reilly and Daehler, 1986; Ivanov-Kholodny and Mikhailov, 1986]. A possible

way to implement such diagnostics is thought to be through measuring the characteristics of a system of high-latitude radio paths in the decameter band. The objects to be diagnosed are the auroral substorms, magnetic storms, main ionization trough, daytime polar cusp, sporadic layers of the ionosphere, such as E_s and others. The said geophysical phenomena have significant effect on decameter radio wave propagation patterns [Cannon, 1989; Blagoveshchenskij, 1981; Besprosvannaya and Lukashkin, 1990]. If this pattern is thoroughly studied, given the data on the radio waves propagation conditions at a specified moment of time, one can form a general idea of the geophysical phenomena and predict their later evolution.

In section 2 of the present paper the feasibility of diagnostics of the auroral substorms are discussed. The results of Blagoveshchenskij and Zherebtsov [1987] are extended to a system of oblique sounding of paths of the ionosphere. The system comprises two subauroral radio paths, one transauroral path, one auroral-polar link, and one polar path. The substorm effect is shown to manifest itself differently on all the paths. In section 3 the principles of the diagnostics of the main ionization trough are considered. Here the oblique sounding of iono-

Copyright 1992 by the American Geophysical Union.

Paper number 91RS02632.
0048-6604/92/91RS-0263\$08.00

TABLE 1. Experimental Path Characteristics

| Path | Length, km | Corrected Geomagnetic Coordinates of the Wave Reflection Point, deg | |
|--|---------------|---|------------|
| | | ϕ' | λ' |
| 1.1 Moscow-Yugorsky Shar (subauroral) | 1900 | 57.6 | 122.9 |
| 1.2 Moscow-Dixon (subauroral) | 2700 | 59.6 | 131.5 |
| 1.3 Moscow-Heiss Island (transauroral) | 2850 | 62.9 | 122.6 |
| 1.4 Heiss Island-Yugorsky Shar (auroral polar) | 1200 | 69.0 | 139.9 |
| 1.5 Heiss Island-North Pole 22 station (polar) | 1900 | 75.2 | 179.6 |
| 2.1 Tiksi Bay-Norilsk (auroral) | 1450 | 65.1 | 179.3 |
| 2.2 Dixon-Norilsk (auroral) | 500 | 65.9 | 157.1 |

sphere is not effective enough, so preference is given to analysis of the statistics of the signal amplitudes at the output of the radio channel. Section 4 deals with the diagnostics of the daytime polar cusp. In the absence of data on the signal statistics, close attention is paid to the oblique soundings as a cusp diagnostic means at the northernmost path situated in the polar cap.

2. AURORAL SUBSTORMS

The nonstationarity of the polar and subpolar ionosphere is caused by the effects of various perturbations which include substorms. During a substorm there are noticeable changes of electron concentration in the bulk of the ionosphere. These changes are usually analyzed using data either of vertical sounding or the incoherent scatter technique. The present work employs the data of the oblique sounding of the ionosphere. The experimental radio paths are shown in Table 1.

As shown in Figure 1, the reflection points of the paths are located in different areas of large-scale, high-latitude structures. Paths 1.1, 1.2, and 1.3 are influenced by the main ionization trough, paths 1.4, 2.1, and 2.2 by the auroral oval. Path 1.5 is located in the polar cap. Single nighttime substorms with a sharp breakup against an unperturbed background

have been studied. The ionograms of the oblique sounding of the ionosphere were used to determine parameters of radio wave propagation: the maximum observed frequencies (MOF) of reflection from the F_2 and E_s layers of the ionosphere (F_2 MOF and E_s MOF), the lowest observed frequencies (LOF). The data used were for 1978–1979 and 1982.

The oblique sounding ionograms data were subdivided into geophysically homogeneous groups according to seasons and time of the day. A graph of temporal variations of the radio wave propagation parameters was plotted for each substorm and each path. This is illustrated in Figure 2, where the vertical line designates the start of the sharp evolution of a baylike perturbation (breakup).

Figure 3 gives as an illustration some curves for the F_2 MOF parameter for winter 1982 during substorms for the Moscow-Dixon path (1.2). The vertical line shows the moment of the substorm breakup. Values of F_2 MOF can be seen to differ for each substorm. However, the dynamics of the changes due to the substorm manifestation are about the same within the limits of each event. In the initial phase of a substorm an increase of F_2 MOF is observed, then in the active phase there is a drop and later the return to the initial level.

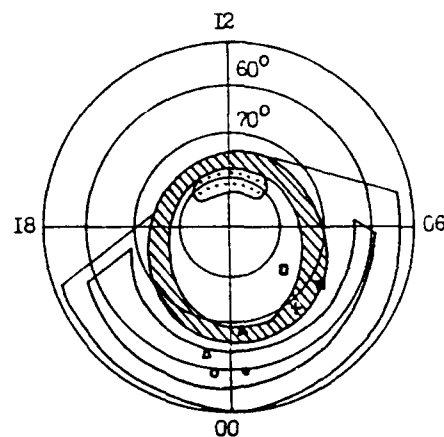


Fig. 1. The position of the mid points of oblique sounding paths in relation to the large-scale inhomogeneities of the F region of ionosphere. (open circle) Moscow-Yugorsky Shar (1.1), (closed circle) Moscow-Dixon (1.2), (open triangle) Moscow-Heiss Island (1.3), (closed triangle) Heiss Island-Yugorsky Shar (1.4), (open square) Heiss Island-North Pole 22 st. (1.5), (closed square) Tiksi Bay-Norilsk (2.1), (cross) Dixon-Norilsk (2.2), (diagonal pattern) auroral zone, (solid line) main ionospheric trough, and (dotted pattern) daytime polar cusp.

To reveal general regularities of the variations of the wave propagation parameters during substorms, the data were averaged and relative characteristics ($\Delta\text{MOF} \%$ and $\Delta\text{LOF} \%$ referred to the median value) were introduced. An example of such averaged curves is given in Figure 4. In the left-hand part (designated as "a") there are the curves $\Delta F_2 \text{ MOF}$ obtained by averaging the data for substorms with intensity of $A = 250\text{--}750 \text{ nT}$ and the curves in the right-hand part (designated as "b") correspond to the substorms with $A > 1000 \text{ nT}$. The graphs

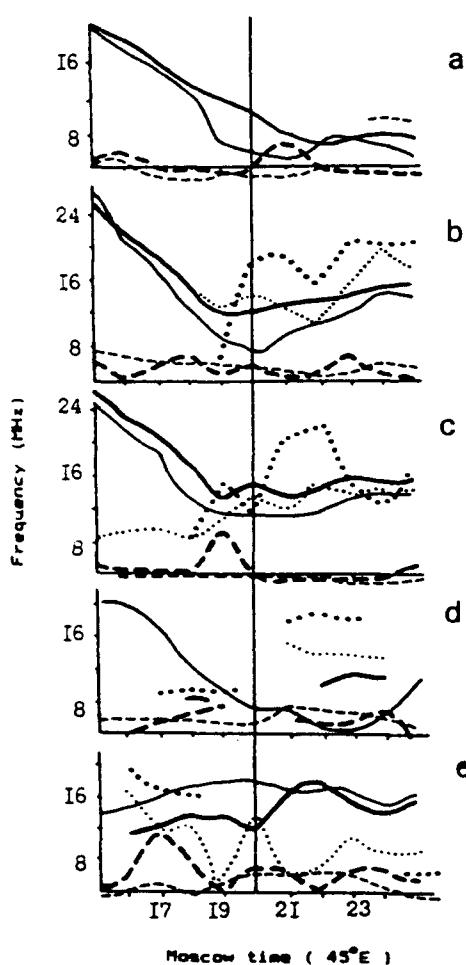


Fig. 2. The substorm effect on propagation characteristics from the data of oblique sounding path network observed on December 12, 1978. Paths: (a) Moscow-Yugorsky Shar; (b) Moscow-Dixon; (c) Moscow-Heiss Island; (d) Heiss Island-North Pole 22 station; (e) Heiss Island-Yugorsky Shar. (Dark solid line) $F_2 \text{ MOF}$, (light solid line) median $F_2 \text{ MOF}$, (dark dashed line) LOF, (light dashed line) median LOF, (dark dotted line) $E \text{ MOF}$, (light dotted line) median $E \text{ MOF}$.

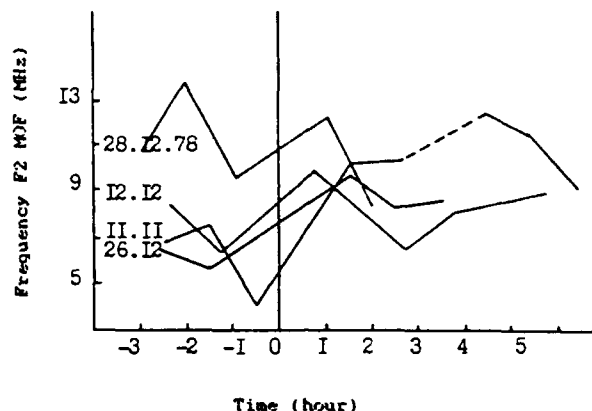


Fig. 3. Variations of the $F_2 \text{ MOF}$ parameter during winter substorm at the Moscow-Dixon path.

show that during a substorm two effects are observed: a general increase of the level and a negative bay in the values of the $F_2 \text{ MOF}$. Here the bay practically disappears from winter to summer in less intensive substorms but can be traced consistently in all seasons for the case of strong substorms. The evolution of the baylike component of $\Delta F_2 \text{ MOF} \%$, which is overlaid by a negative bay, begins before the breakup of a substorm in the magnetic field. The maximum of the $\Delta F_2 \text{ MOF} \%$ is, as a rule, found at the moment the AE index curve maximizes.

In the average over all the data the main features of the $F_2 \text{ MOF}$ variation during a substorm at subauroral paths 1.1, 1.2, and transauroral 1.3 are an increase of electron concentration N_e relative to the median ($\Delta F_2 \text{ MOF} > 0$) several hours before the beginning of the bay, a decrease of concentration in the period of explosive phase ($\Delta F_2 \text{ MOF} < 0$) and again an increase of N_e at the end of the bay ($\Delta F_2 \text{ MOF} > 0$) for several hours with a gradual weakening of the effects back to the normal state. For the path 1.4 in the auroral zone the reflections from the F_2 layer are irregular and no regularities are revealed here. During a substorm on path 1.5 the values of the $F_2 \text{ MOF}$, on the whole, oscillate about the median for the period of approximately ± 5 hours relative to the bay maximum without noticeable regularities.

The values of the $E_s \text{ MOF}$ at the subauroral paths are in general less than the median at initial stage of the substorm evolution ($\Delta E_s \text{ MOF} < 0$). Later there is a rise in the bay maximum vicinity ($\Delta E_s \text{ MOF} >$

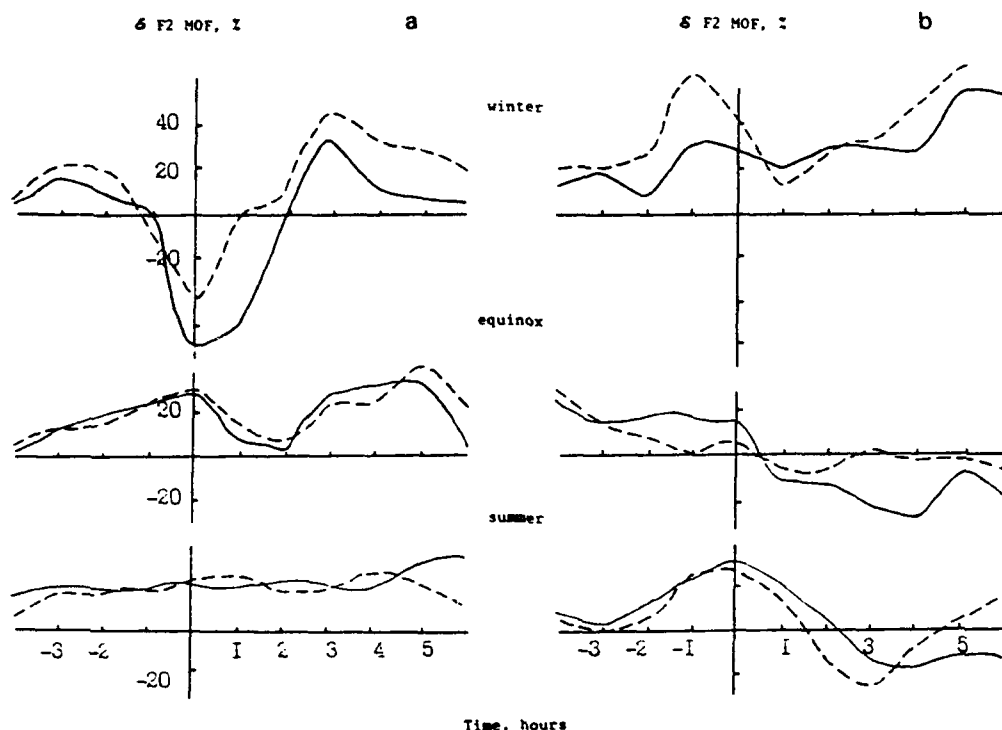


Fig. 4. Variations of the ΔF_2 MOF parameter during substorm by averaged data; (a) weak substorms; (b) strong substorms; (solid line) path Moscow-Dixon. (dashed line) path Moscow-Heiss Island.

0) and again a decrease relative to the median (ΔE , MOF < 0) for several hours. At the auroral path and the path in the polar cap the variations pattern of E , MOF is not stable.

The LOF variations reduce mainly to a growth of the values relative to the median several hours before the beginning of the substorm evolution ($\Delta LOF > 0$), a fall off during the bay ($\Delta LOF < 0$) and again a rise at the final stage of the perturbation ($\Delta LOF > 0$) for several hours as well.

Substorm diagnostics using HF paths can furthermore be based on acquisition and analysis of a set of signal statistical parameters whose variations in different phases of the substorm have certain regularities. This is illustrated in Figure 5. Here \bar{X} is the mean signal level in relative units, m is the Nakagami law parameter characterizing the depth of signal fading [Nakagami, 1960], τ_k is the temporal correlation radius characterizing the rate of rapid signal fading.

During the evening and night hours on radio paths 2.1 and 2.2 which cross the auroral zone, at the signal reception site in Norilsk, there is an increase of values of \bar{X} , m , τ_k , and other parameters several

hours before the breakup begins ($T = 0$). About an hour before the breakup a sharp drop of the parameters begins. In the breakup phase the parameters values are lower than before and after breakup at all the paths, subauroral ones included. In the recovery phase an increase of the said parameters is observed as well at all the auroral and subauroral paths of length up to 3000 km.

Considering the substorm effects on all parameters of radio wave propagation, the oblique sounding parameters included, one can make the following conclusion. In the active phase of a substorm (breakup) ionization is redistributed throughout the ionosphere. At the ionospheric F_2 layer the electron concentration becomes less and at the level of the layers E and D correspondingly greater.

3. THE MAIN IONIZATION TROUGH

The trough has a substantial effect on HF signal passage especially at the wave reflection point. It causes sharp changes in the signal level and structure as well as the azimuthal angles of wave arrival.

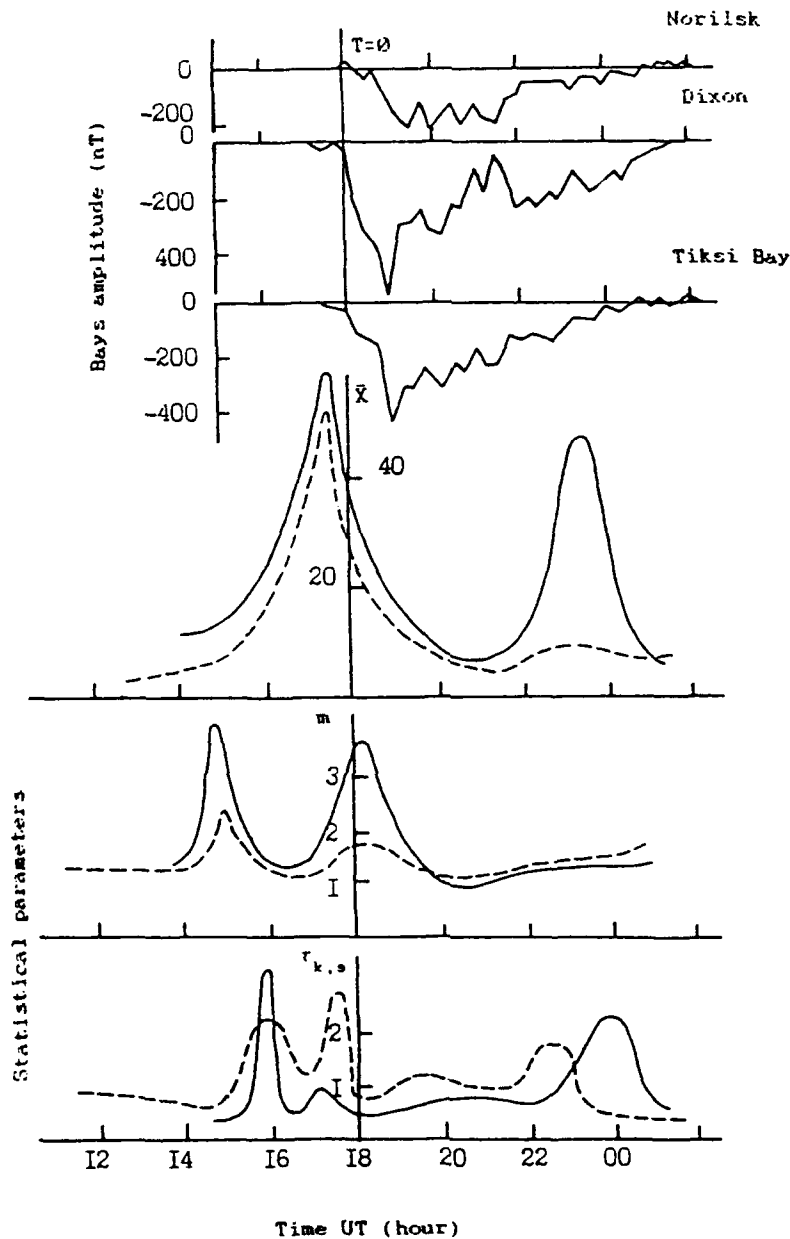


Fig. 5. Variation of signal parameters \bar{X} (mean value), m (Nakagami parameter), and τ_k (temporal correlation radius) at Norilsk-Tiksi Bay (solid line curve) and Norilsk-Dixon (dotted line curve) paths for the period of isolated substorm on December 28, 1973.

Most radio paths located at the latitudes $\Phi' = 50^\circ - 70^\circ$ or crossing them are in this sense under the influence of the trough. Figure 6, for example, illustrates the regularities of the changes of some statistical signal characteristics on path 2.2

(Dixon-Norilsk) at the moment of transition of the path reflection area from the trough to its northern wall. The moment is marked at the figure by an arrow. Here β^2 , the ionosphere turbidity factor, is defined as the ratio of energies of mirror signal

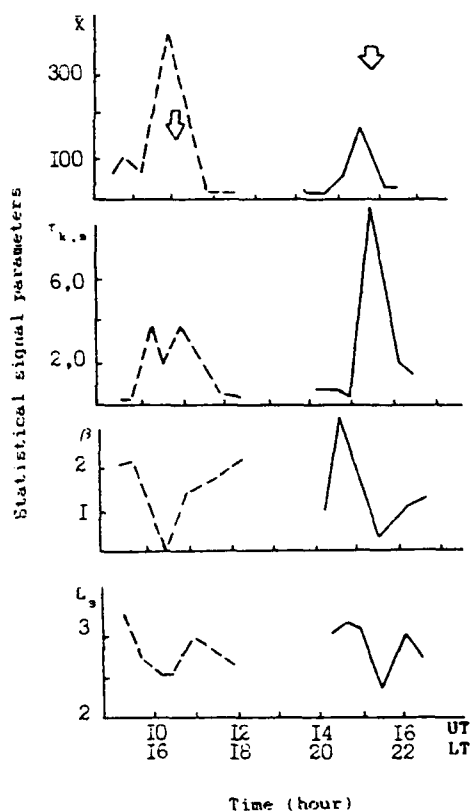


Fig. 6. Signal parameters fluctuations near the trough northern boundary (arrows) at the Dixon-Norilsk path across the auroral zone. (solid line) $Kp < 3$, (dashed line) $Kp \geq 3$.

component and the sum of scattered ones. L_3 is the nonstationarity parameter: $L_3 = \int_0^T |K(\tau)| d\tau / \int_0^T K^2(\tau) d\tau$, where T is the selective process duration and $K(\tau)$ is the standardized correlation function of this selection. The process being quasi-stationary under the condition $2 < L_3 < 3$. Figure 6 illustrates substantial changes of the parameter values.

4. THE DAYTIME POLAR CUSP

The daytime polar cusp diagnostics based on propagation conditions may be carried out by the ionosphere oblique sounding method. The specifics of the influence of the complex structure of the auroral ionosphere in the area of the daytime polar cusp on the passage of the HF signal is traced in an example of polar path of oblique sounding (1.5 Heiss Island—North Pole 22 station) during 1979. The middle point of the path at the time was

situated at the latitude $\Phi' = 75^\circ$. The cusp in the perturbed conditions $Kp > 3$ is, on the average, also situated at this latitude. The hourly ionograms of the oblique sounding were used to determine the maximum observed frequency (F_2 MOF) as the most informative parameter of the propagation conditions and the general pattern of the signal multipath structure. All experimental material was divided into three seasons, where in its turn a classification was made according to the degree of magnetic disturbance characterized by the index Kp ($Kp = 0 - 1$, $Kp = 2$, $Kp \geq 3$). The autocorrelation coefficient $r(\tau)$ with a shift of $\tau = 1$ h was calculated using the set of hourly values of the F_2 MOF for each season over a 24-hour period. It is an indicator of the nonstationarity of the radio channel state: the less $r(\tau)$, the higher the instability and more complex the ionosphere dynamics at the radio wave reflection point. Figure 7 shows the 24-hour cycle of $r(\tau)$ variation for $Kp \geq 3$ in different seasons. One can see that during the period there are four distinct maxima in the values of $r(\tau)$ which are given in the figure as dark sectors. The daytime cusp position (an asterisk in Figure 8) is manifested in the minimum of $r(\tau)$ values respectively at 1400 LT in summer, 1300 LT in equinox, and 1200 LT in winter which is an evidence of the ionosphere nonstationarity in the cusp area. The minimum at 1700 LT (winter), 1800 LT (equinox), and 2000 LT (summer) is due to the transition of the path reflection point from the auroral oval to the polar cap (triangles in Figure 8) and the soft precipitation zone boundary near the latitude $\Phi' = 75^\circ$. The minimum at 0000–0400 LT can be explained by the inhomogeneous structure of the ionosphere caused by the ionization peaks manifesting the strongest way at 1800 LT. The described regularities are also traced in the features of the mode structure of the HF signal. The detailed analysis of all scopes of the phenomena has shown that the cusp effects are in the main manifested in the emergence of additional anomalous modes with small intermodal delays. The examples of ionograms of ionosphere oblique sounding are given in Figure 9. The presence of additional modes indicates that with a cusp area emerging at the path there are inhomogeneities of a scale allowing realization of several (twin) trajectories spaced at the middle points by ≈ 200 – 300 km. These effects manifest themselves most strongly in the winter period.

Thus the suggested methods of diagnostics and

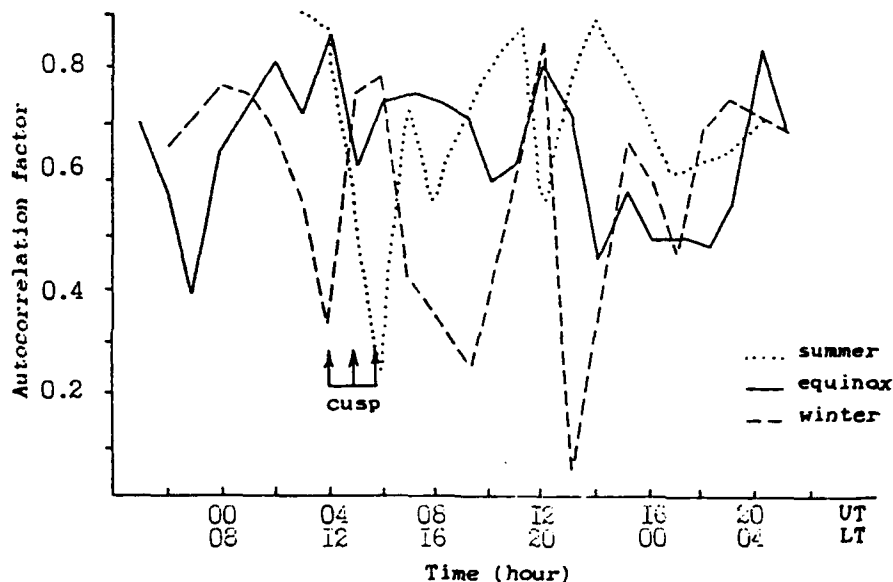


Fig. 7. The 24-hour variation of the autocorrelation factor $r_1(\tau)$ with 1-hour shift for the F_2 MOF values (summer, equinox, winter).

consequently real-time prediction of auroral substorms, main ionization trough, and daytime polar cusp by the characteristics of propagating radio waves have important ramifications for problems of propagation and their applications.

5. CONCLUSION

It has been shown in the present paper that large-scale inhomogeneities of the medium and dynamically evolving processes in high-latitude ionosphere are the cause of the wide range of radio

channel characteristics. However, the geophysical manifestations on each particular radio path have certain statistical regularities which allow us to describe the radio channel models and to relate variations of the propagation parameters to some geophysical factor (substorm, main ionization trough, daytime cusp). Having at one's disposal a network of radio paths with characteristic regularities of signal parameters variations, one can proceed to solving two problems. First, the short-time prediction of high-latitude effects by typical changes of propagation parameters on the paths in

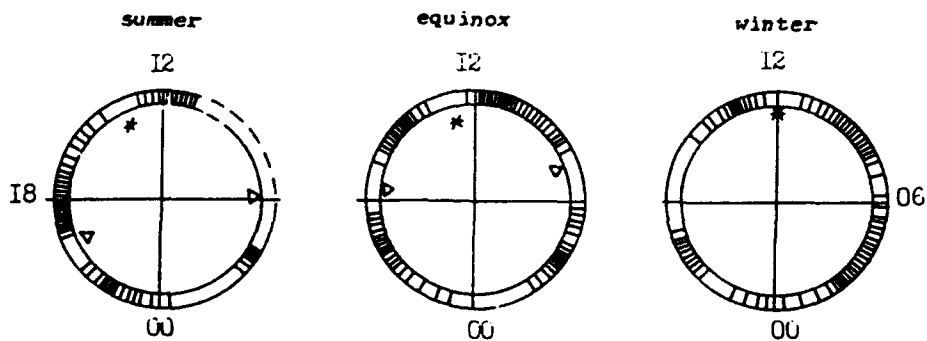


Fig. 8. Graphs of maxima (darkened) and minima (undarkened) of the factor $r_1(\tau)$ values as a result of influence of different structures of polar ionosphere on the reflection point of a radio path. The asterisk signifies a cusp position.

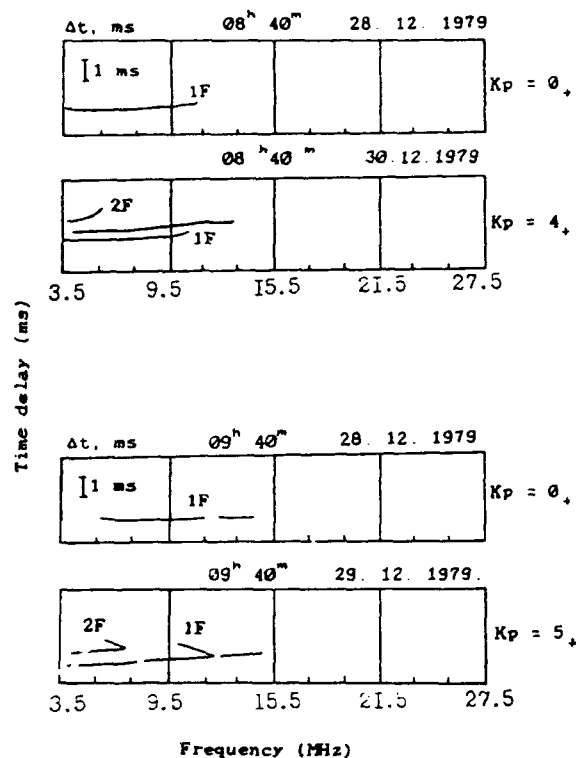


Fig. 9 Examples of ionograms of the oblique sounding of ionosphere: single modal in the absence of a cusp multimodal under the impact of a cusp.

real time. Second, the farther and final prediction of expected changes of radio wave propagation conditions under the impact of these prognostic factors.

The basis of methods of qualitative short-time (up to several hours) prediction of ionospheric propagation paths is the availability of radio signal's model and channel diagnostics (the real-time acquisition of working information on the structure or the parameters of radio signals traveling along the propagation paths under study). The working information goes at once to the data processing center. The information should be either a set of values of the mean level of signals arriving simultaneously from several radio transmitters over channels (paths) under study or one signal over one path but necessarily with on-line evaluation of the set of its statistical characteristics: the Nakagami parameter

m , the temporal correlation radius τ_k , the mean signal amplitude \bar{X} , the nonstationarity parameters. If oblique sounding of the ionosphere is carried out, a set of propagation characteristics is taken: MOF, MUF, LOF, signal time delay, and so on. The maximum volume of diagnostic data results when these observational methods are used together on several paths simultaneously. In other words, the prediction of radio channel characteristics should be carried out with a set of informative signs determined by the history (learning sample) since each sign by itself is not sufficient for prediction. Their combining into a certain set gives much greater information.

REFERENCES

- Akasofu, S. I., *Polar and Magnetospheric Substorms*, D. Reidel, Hingham, Mass., 1968.
- Besprozvannaya, A. S., and V. M. Lukashkin, Dynamics of the high-latitude ionosphere and variation of HF radio wave propagation conditions during world storms, *Geomagn. Aeron.*, 30, 682-685, 1990.
- Blagoveshchenskij, D. V., *High-Latitude HF Radio Wave Propagation*, Nauka, Moscow, 1981.
- Blagoveshchenskij, D. V., and G. A. Zherebtsov, *High-Latitude Geophysical Phenomena and Prediction of HF Radio Channels*, Nauka, Moscow, 1987.
- Cannon, P. S., Morphology of the high-latitude ionosphere and its implication for HF communications systems, *IEEE Proc. I.*, 136, 1-10, 1989.
- Eather, R. H., Polar cusp dynamics, in *The Polar Cusp*, edited by J. A. Holtet and A. Egeland, pp. 149-163, D. Reidel, Hingham, Mass., 1985.
- Galperin, Yu. I., L. D. Sivtseva, V. M. Filippov, and V. L. Khalipov, *Subauroral Upper Ionosphere*, Nauka, Novosibirsk, 1990.
- Hunsucker, R. D., Effects of the auroral and polar-cap ionosphere on radio propagation: A mini-review, *IEEE Trans. Antennas Propag.*, in press, 1991.
- Ivanov-Kholodny, G. S., and A. V. Mikhailov, *The Prediction of Ionospheric Conditions*, D. Reidel, Hingham, Mass., 1986.
- Nakagami, M., The m -distribution—A general formula intensity distribution of rapid fading, in *Statistical Method in Radio Wave Propagation*, pp. 3-36, Pergamon, New York, 1960.
- Reilly, M. H., and M. Daehler, Sounder updates for statistical model predictions of maximum usable frequencies on HF sky wave paths, *Radio Sci.*, 21, 1001-1008, 1986.
- D. V. Blagoveshchenskij, The Leningrad Institute of Aviation Instrument Making, 67 Hertsen Str., Leningrad, 190000, USSR.
- L. V. Egorova and V. M. Lukashkin, Department of Geophysics, The Arctic and Antarctic Scientific Research Institute, 38 Bering Str., Leningrad, 199226, USSR.

Length measurements of mid-latitude scintillation irregularities

John W. MacDougall

Department of Electrical Engineering, University Western Ontario, London, Ontario, Canada

(Received January 30, 1991; revised September 18, 1991; accepted September 24, 1991.)

The lengths of irregularities which produce 150-MHz amplitude scintillations have been measured at 43°N, 81°W (geographic) using arrays of receivers with large spacings. The average length (major axis radius) of the irregularities was 6.1 km. This is much shorter than expected and implies that the measurements are of "young" irregularities, less than 1 minute old. These irregularities appear to be a large, 25-50% perturbation of the background density.

INTRODUCTION

A previous study [MacDougall, 1990] using spaced receivers with spacings of several kilometers to measure the length of the ionospheric irregularities which cause scintillations found lengths of 5-10 km along the magnetic field direction. These lengths were substantially longer than those obtained from arrays of receivers with smaller spacings, and the reasons for the discrepancies were discussed [MacDougall, 1990]. At that time it seemed likely that the 5-10 km was still an underestimation of the length and that a repeat of the experiment using arrays of receivers with increasingly large spacings would yield measurements which were much longer.

The basic reason for expecting much longer lengths is that the irregularities, which cause scintillations, have cross-section dimensions of the order of hundreds of meters, and irregularities of such dimensions have lifetimes of the order of hours [Vickrey and Kelley, 1982]. Thus any irregularity which is randomly observed would usually have had adequate time to diffuse along the Earth's magnetic field and would appear as just a field-aligned percentage depletion or enhancement of the background ionization. It was therefore expected that correct measurement of the lengths would give values which would be of the order of the vertical scale size of the ionosphere.

The present experiment set out to verify that the lengths were very long by using receiver arrays with larger spacings than in the previous experiment. Initially, a spacing of 19.3 km at an azimuth of 35° was used. For this spacing the amplitude scintillation records from the two receivers were cross correlated, and it was expected that a substantial maximum in the correlation functions would occur at zero

time lag during the portion of the pass when the long irregularities simultaneously projected onto the two stations. The correlation maximum should shift in time lag in a predictable manner as the satellites passed by the observing sites. However, the measured cross-correlation functions showed only the random maxima one would expect from any two data samples which are uncorrelated but have similar periodicities.

The quasi-periodic nature of amplitude scintillations is due to the Fresnel band-pass filtering process involved in producing these amplitude scintillations. The correlation level of the randomly occurring maxima was typically 0.25 for the sample lengths used. If the irregularity lengths, as projected on the ground, had been 19.3 km or greater the expected correlation maxima would have been greater than 0.5 and would have stood clearly above the random 0.25 maxima. Therefore the ground-pattern length of the scintillations had to be shorter than 19.3 km. The ionospheric lengths which would give a 19.3-km projection along the line joining the two stations depended primarily on the elevation angle of the satellite when it was located so as to give the correct projection. It was possible to select passes for which the ionospheric lengths would be either shorter or longer than 19.3 km. However, for ionospheric lengths much shorter than 19.3 km the elevation angles of the satellites would be very low.

A variety of passes were examined and for only one pass could a significant cross correlation between the station pair be detected. Conditions during this pass were indistinguishable from conditions of a number of other similar passes for which no cross correlation could be detected.

Following this initial largely unsuccessful length measurement, several arrangements of the receivers with shorter spacings were tried, and the results of these measurements will be presented in the next section.

Copyright 1992 by the American Geophysical Union

Paper number 91RS02917.
0048-6604/92/91RS-02917\$08.00

LENGTH MEASUREMENTS

The measurements used the Navy Navigation Satellite System (NNSS) series of satellites and recorded the amplitude scintillations at 150 MHz. The NNSS satellites are at heights near 1100 km in polar orbits. They therefore pass by the observing site in approximately a north-south direction. Figure 1 shows a sketch of the experimental situation. It may be seen that as the satellite moves toward north or south the projection of the field-aligned irregularity on the ground moves in the opposite direction with a velocity which is determined by the relative geometry of irregularity and satellite. The velocities of the two ends of the ground pattern are slightly different, due to the geometry, so that the ground projection rotates slowly in azimuth as the satellite passes. The length of the ground projection also changes, being at a minimum when the satellite is approximately at the observer's latitude for a station pair with east-west spacing.

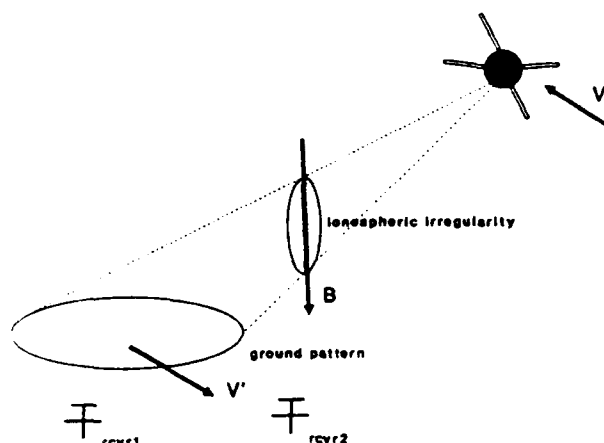


Fig. 1. Showing the general geometry of the measurements. The radio beacon signal from the satellite gives a diffraction pattern of the field aligned irregularity on the ground, where it is observed by spaced receivers. The satellite motion causes the ground pattern to move as shown.

Several station arrangements were used during these measurements. For most of the measurements the stations were located approximately east-west of one another so as to have the projection of the irregularities simultaneously over the station pair when the satellite was near the observer's latitude. As the satellite moves away from this position, the geometry becomes less optimum for determining the length of the irregularities. This will be shown in more detail later in the paper. The stations were all located in the vicinity of London, Ontario, Canada (43°N, 81° W geographic, 54° geomagnetic latitude).

Correlations used 10 s of data. Because of the changing geometry of satellite and receivers, the time shift between

seeing the projection of the same irregularity on the two receivers changes as the satellite passes. This may be visualized from Figure 1. For 10-s samples of data this change in time shift is small unless the satellite passes by close to overhead (i.e., the satellite longitude is approximately the same as the observer's longitude). It is possible to put in a correction factor to the time lags when doing the correlations, so as to account for the change in time shift, but for this correction factor it is necessary to assume a specific height for all the irregularities. Since the irregularities are usually spread over a range of heights [MacDougall, 1981], using one specific height for all the irregularities is not a good assumption. Therefore the correction factor was not used except for the near-overhead passes for which the change in time shift becomes appreciable during the 10-s samples. For these corrections the irregularities were assumed to be at 350-km height.

Restricting the sample length to 10 s meant that random cross-correlation maxima observed for time lags when there could be no possibility of seeing the same irregularity projection at the two receivers were of the order of 0.25. Using longer samples reduced the amplitude of these random correlation maxima. The amplitude of any real correlation maxima did not change noticeably for typical data sets when the sample lengths were increased beyond 10 s, so most of the analysis divided the records into 10-s samples.

Amplitude scintillations were recorded at 50 samples per second, which is approximately equivalent to one measurement each 50 m of the passage of the line of sight to the satellite through the F region ionosphere. Receiver time constants were 5 ms so the 50 samples per second were largely independent.

Figure 2 shows the results for one satellite pass. The amplitude of the cross-correlation maxima between stations spaced 4.6 km approximately east-west is shown by the dashed curve. From these cross-correlation amplitudes the lengths of the ground patterns may easily be determined as follows: The cross correlation is assumed to be a Gaussian, $\rho = \exp(-(\chi/c)^2)$, where χ is a distance. Use the measured cross-correlation maximum amplitudes and $\chi = 4.6$ km to solve for c , then solve for what length, $\chi_{1/2}$, would make $\rho = 1/2$. $\chi_{1/2}$ is the length (one side) along the major axis of the ground pattern to a correlation of 1/2 and is the "length" definition used in this paper.

The $\chi_{1/2}$ lengths of the ground patterns can be used to determine the ionospheric lengths by taking into account the geometry of the ray paths to the satellite for the sample interval. The ionospheric irregularities are assumed to be field aligned and at a height of 350 km for this calculation. The solid curve in Figure 2 shows the ionospheric lengths corresponding to the correlations shown by the dashed curve in this figure.

In calculating the lengths of the ground patterns an alternative shape for the correlation was to assume the part

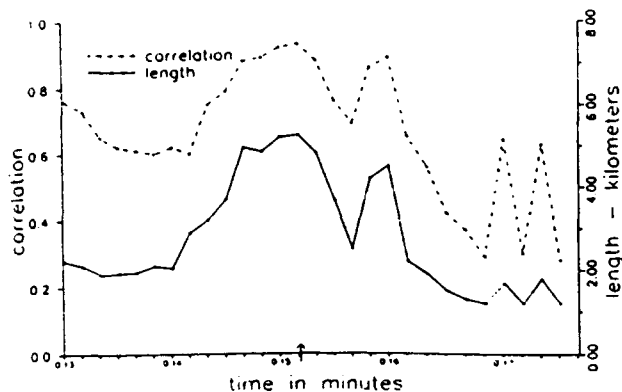


Fig. 2. Maximum cross correlation (dashed) and calculated irregularity length, in the ionosphere, (solid) for one satellite pass on April 21, 1986. The center of the pass was at about 0015 UT, as shown on the time scale. The station pair for the correlation was spaced 4.6 km approximately east-west.

near zero timeshift is a cosine function. For this function the calculated ionospheric lengths were slightly shorter than using the Gaussian function. In most cases the difference in calculated length was less than 1 km.

The observations which will be shown next used three stations located along an approximately east-west line. The second station was 310 m to the west of the first station, and the third station was a further 4.3 km to the west. Since the maximum cross correlation should be almost perfect between the two closely spaced stations, if there was no random noise in the data, the less than perfect cross correlations which were actually measured were used to correct the length calculations for random noise effects. This procedure is similar to that discussed by Meek [1990].

In Figure 2 the correlations and hence the calculated lengths decrease as the satellite moves away from the time when the irregularity projected along the line of stations. This time is marked by an arrow on the time scale and will be referred to as t_0 . One might expect that the correlations would increase rather than decrease as one moves away from t_0 since for earlier and later times the ground projection of the irregularities will be longer and hence the correlation between stations separated 4.6 km should increase. This does not happen for three reasons. The first effect is: the station pairs are separated approximately east-west, and at times earlier and later than t_0 the irregularities project at an increasing angular rotation with respect to the east-west line. This rotation of the ground pattern means that the effective station spacing is increased when the ground pattern moves across the two stations in a north-south direction as a result of satellite motion. Figure 3 illustrates this effect. The increase in effective spacing will approximately cancel the effects of longer ground projection of the irregularities at earlier and later times.

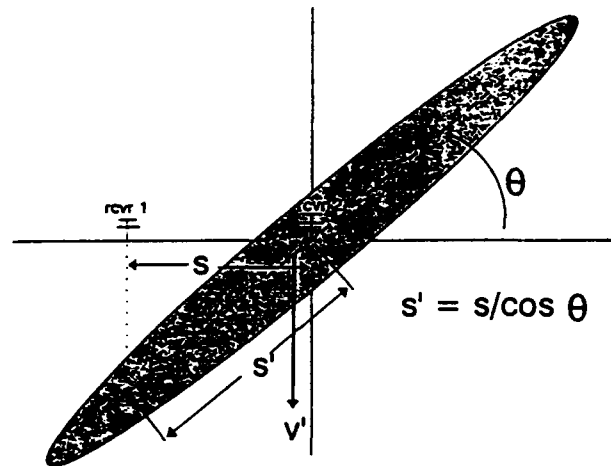


Fig. 3. Illustrating the geometrical effects of a ground pattern rotated with respect to the line joining the station pair.

Taking this increase in effective spacing into account, one might therefore expect the correlation values to stay approximately constant during the pass. This does not happen because of a second effect.

This second effect is also geometrical and is illustrated by Figure 4. This figure shows an exaggerated picture of the location of scintillations which might be viewed from two receiving sites at two different satellite positions. The two positions "A" and "B" are two positions along the satellite north-south trajectory. Observed scintillations might be considered to be a summation of effects due to the irregularities which are contained in the two

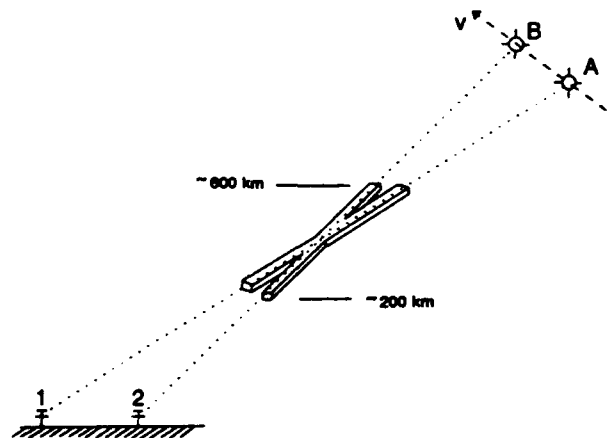


Fig. 4. Illustrating the nonoverlap of the ionospheric regions containing scintillations when the satellite is at two different positions. The regions are such that their approximate center is projected to station 1 when the satellite is at A and to 2 when the satellite is at B.

(intersecting) boxes. The top of the boxes is arbitrarily taken as 600 km which is typical of the upper heights when one measures scintillation irregularity heights [MacDougall, 1981]. Similarly, 200 km is taken as the lower height. The two boxes overlap around the peak of the F region, which is the most common height when measuring scintillation irregularity heights. Since the two boxes only partially overlap, one might expect the cross correlation would be reduced, since the irregularities in the nonoverlapping parts of the two boxes, while contributing to the scintillations, would not be correlated.

Calculations were done to see how much the correlations might be reduced by the nonoverlap of the regions being observed. The irregularities were assumed to be distributed in height with a Gaussian distribution peaking at 400 km and decreasing to $1/2$ the peak value at approximately 225 and 575 km (based on MacDougall [1981]). The correlation was assumed to be reduced by the weighted fraction of irregularities not in the overlapping parts of the boxes. The thickness and width (assumed to have equal dimensions) for each box should be comparable with the Fresnel zone size. A width of 1 km gave decreases of correlation comparable with those observed. (One km is approximately the Fresnel wavelength, $(2\lambda z)^{1/2}$, for these measurements.) The calculation also showed that the correlation decrease should be larger for near overhead passes, and this was seen in the measurements.

If the irregularities are not as widely dispersed in height, the effect of nonoverlapping is much reduced. Since all the measurements showed decreases of correlation similar to that seen in Figure 2, it is inferred that all the measurements are of scintillation patches in which the irregularities are widely distributed in height.

A third effect which might reduce the correlation at times earlier and later than t_0 is caused by the time delay in seeing the ground pattern at the two receivers (see Figure 3). If there are rapid changes in the irregularities, then the correlation would be reduced by these changes if the irregularities are viewed at two slightly different times. In the Full Correlation Analysis (FCA) developed by Briggs *et al.* [1950] and Phillips and Spencer [1955], these random changes are ascribed to a velocity component. This velocity component would also be affected by other factors which reduce the correlation, such as those noted above.

Because all these effects vanish at t_0 at which time the line of receivers should be viewing exactly the same group of irregularities (at slightly different heights) with no time delay and no orientation angle effects, the lengths which will be shown later are those determined from the average correlations observed around t_0 on each pass. This gives one length measurement per pass, which greatly reduces the number of length determinations but increases confidence that the measurements are meaningful.

These measurements are shown in Figure 5. Here the

ionospheric length determined at t_0 for each pass is plotted versus the longitude of the pass at t_0 . Although there is a scatter of values, there is no obvious functional relationship between the lengths and the longitude. This lack of a functional relationship is somewhat reassuring that the lengths shown actually represent the ionospheric lengths. This is because the lengths measured on the ground should be approximately a projection of the ionospheric lengths, and the ground lengths varied greatly with satellite longitude. Ground lengths were less than 1 km for the near overhead pass, and 10-15 km for the passes at the low-elevation angles seen on Figure 5 around -65° and around -95° longitude.

The average ionospheric length for the data shown on Figure 5 is 6.1 km. The intensity of the scintillations for the individual Figure 5 data points varied from quite weak ($S_4 < 0.05$) to moderate ($S_4 \approx 0.35$).

Since the 6.1-km length was much shorter than expected, some of the individual data recordings were examined to see if there was evidence that the "individual" irregularities were as short as this or whether this result was due to some sort of statistical averaging.

Figure 6 shows a portion of one sample of data in which a particularly well-defined amplitude fluctuation may be seen on the recording of the signal amplitude from receivers 0,1,2 (receiver locations are indicated on the figure) but is not seen on receiver 3, which was 4.6 km distant. Thus this "individual" irregularity projected with a length on the ground of less than 4.6 km. This record is from a near overhead pass, and the ground length calculated from the cross correlation function for the data sample which includes this time was approximately $3/4$ km. Therefore the length measured from the correlation functions ($3/4$ km for this example) is compatible with the

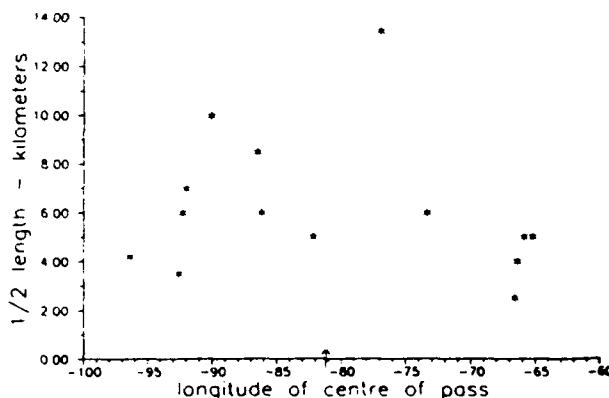


Fig. 5. Length of irregularities measured on a number of satellite passes. The passes are plotted as a function of their longitude position when at the observers latitude. The longitude of the observer is shown by an arrow. The mean length is 6.1 km.

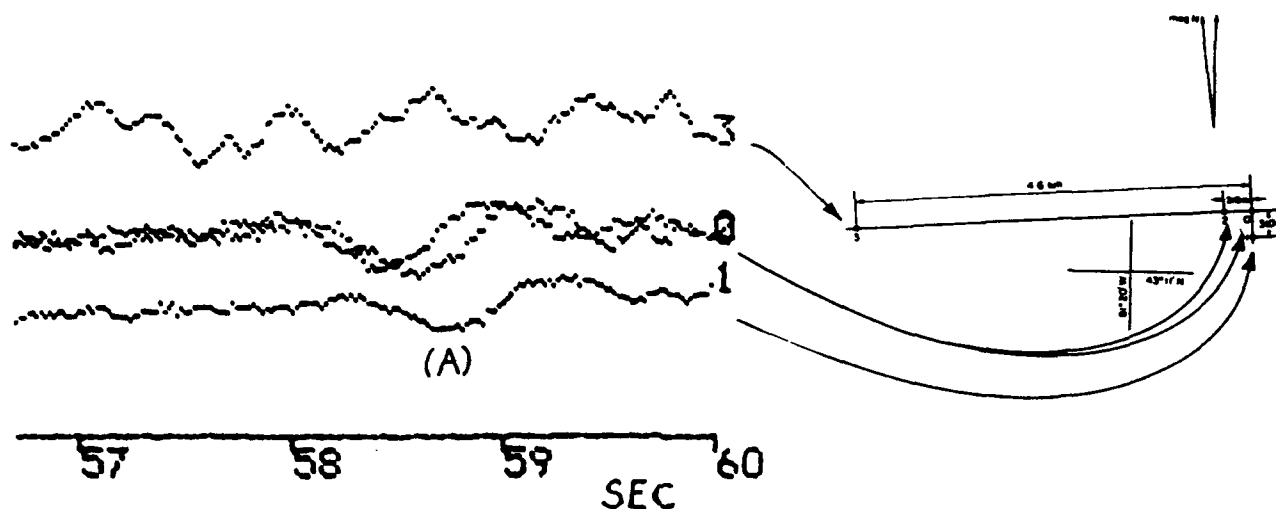


Fig. 6. Details of amplitude scintillations for one near overhead pass showing an amplitude fluctuation (A) seen on receivers 0,1,2 but not on receiver 3, which is spaced 4.6 km distant.

scale sizes of the individual features which may be seen on the spaced receiver records.

The 6.1-km average ionospheric length for the irregularities does not mean that the axial ratios of the irregularities is small. The average minor axis length of the ground patterns (to a correlation of $1/2$) during these measurements was 350 m. This is an average of the minor axis lengths measured using correlation analysis (FCA). Individual lengths varied between 180 and 550 m. The ground minor axis lengths would project to shorter lengths in the ionosphere by the ratio of $1-(h_i/h_s)$, where h_i is the irregularity height and h_s is the satellite height. Taking a typical irregularity height of 350 km and satellite height of 1100 km the projection factor is 0.68, which gives an ionospheric minor axis length of 0.24 km. The average irregularity axial ratio would therefore be 25. An axial ratio of 25 means there would be an appreciable scintillation enhancement if the line of sight to the satellite comes within a few degrees of being aligned along the magnetic field (see Figure 10 of Briggs and Parkin [1963]). Using the formulas given by Briggs and Parkin, the peak magnitude of the enhancement for the irregularities and conditions in this experiment was found to be approximately 4x. The data set contained four passes which had longitudes within a degree of that required for field alignment, and all four showed enhancements when closest to field alignment. Unfortunately, although the magnitudes of the enhancements appeared to be comparable with a 4x enhancement, two of the passes had receiver saturation during the enhancement and the other two were on highly structured scintillation boundaries so that for none of the four passes could a reliable number be obtained for the amount of enhancement.

DISCUSSION

As stated in the introduction, lengths much greater than 6.1 km were expected. To confirm this, a computer model was used to see how fast F region irregularities would lengthen by diffusion. The results of this modeling are shown in Figure 7.

The computer model divides the ionosphere into a grid of vertical and horizontal blocks. There is assumed to be

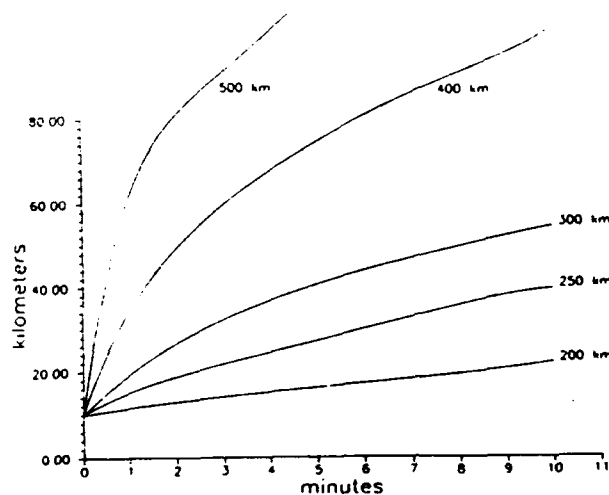


Fig. 7. Length of irregularities versus time as determined from a two dimensional ionospheric computer model. The model ionosphere has 10-km triangular irregularities superimposed at $t=0$. The simulation was done for irregularities at various F region heights, as marked on the curves.

no variation in the third (horizontal) direction. Typical values for temperatures, composition, etc., from ionospheric models are used. The simulation started with typical nighttime ionospheric electron densities. For each block the electron density changes are calculated as a function of time using small time steps and including production, recombination, and vertical and horizontal diffusion.

The model simulations began with an irregularity with a triangular height profile having a length (peak to half amplitude) of 10 km and a peak amplitude 10% of the background density. The curves in Figure 7 all start from the initial length of 10 km. Because the vertical blocks size in the F region part of the model ionosphere were 5 km in the lower part and 10 km in the upper part, a smaller irregularity starting size was not suitable. The horizontal width of the blocks in the model was 100 m for this simulation, and the irregularity had an initial horizontal width of 1 km (10 blocks) with "sharp edges." Horizontal diffusion, as expected, was relatively slow so that, although the sides of the irregularities became less sharp at increasing time during the simulations, the amount of ionization which diffused outside the 10 horizontal blocks was relatively small. Therefore the product of the peak irregularity electron density and the lengths shown in Figure 7 stayed approximately constant. Lengths shown in Figure 7 are for the central axis of the irregularity.

Figure 7 shows that, above the lowest heights in the F region, irregularities diffuse in height extremely rapidly, and by extrapolation, an irregularity of length 6.1 km (as measured) would have to be younger than approximately 1 min. The implication of this result is that if 6.1 km is a valid length measurement, then the irregularities are being detected approximately at the location and time of formation.

This result is somewhat surprising since F region irregularities have lifetimes of many hours [Vickrey and Kelley, 1982] and therefore statistically, any given irregularity would more likely be old than young. A young irregularity is here defined to be one which has not lengthened by more than 6.1 km by diffusion since forming.

It is possible that this bias to young irregularities might be at least partly explained if the younger irregularities were more visible than older irregularities. As noted above, the simulations showed that the product of length times peak electron density stayed approximately constant for the irregularities in the simulation. This meant that the older simulated irregularities had quite a low peak electron density. If irregularities which change the phase of the wave by 180° are assumed to be the most effective, then the integrated content through an irregularity at 150 MHz would need to be 5.6×10^{14} electrons/m². The size of an irregularity in the minor axis direction (minor axis radius) as measured from correlations for this 150-MHz experiment was typically of the order of 250 m, so the peak

electron density perturbation would have been of the order of 1.1×10^{12} electrons/m³. This assumes that an irregularity is spherically symmetrical about its major (field aligned) axis and has a Gaussian distribution as a function of minor-axis radius. It also assumes that the ray path through the irregularity is along the minor axis. The actual ray path is always inclined to the minor axis, typically by the order of 45° for this experiment and this would reduce the peak electron density perturbation required to around 8×10^{11} electrons/m³. The typical peak F region electron density during these measurements was of the order of 2×10^{11} electrons/m³ (Canadian Ionospheric Data for Ottawa, April 1986), so obviously one could not have a depletion of the magnitude calculated. An enhancement of this amplitude is not possible without showing on ionograms as a spread F trace which extends to much higher frequencies than are commonly observed. Therefore, in general, the observed scintillations cannot be due to a single irregularity which introduces a 180° phase change in the ray.

For the cases of near overhead passes the geometry changes so that the ray path through an irregularity is much more oblique. For one pass, which had strong scintillations, the obliquity of the ray would have reduced the peak electron density perturbation requirement to about 8×10^{10} electrons/m³ which is not impossible. Since the ray in its passage through the ionosphere would sample fewer irregularities for a near overhead pass, there seems to be some possibility that for this situation, at least, individual irregularities whose electron density represents a substantial perturbation of the background density are being seen.

The more general case of the scintillations which would be produced by a distribution of many ionospheric irregularities of various sizes has been analyzed by Yeh and Liu [1982], and they show (their Figure 10) that maximum-amplitude scintillations at 125 MHz (the nearest frequency to 150 MHz on their figure) would require rms fluctuations of electron density of greater than 5×10^{10} electrons/m³. This estimate is similar to that for the near-to-overhead pass calculation above, although for their calculation they are considering an ensemble of irregularities having various statistical properties.

Taking the irregularity peak perturbation to be of the order of $5\text{--}10 \times 10^{10}$ electrons/m³ represents a surprisingly large (25-50%) fraction of the peak F region electron density. It would seem that such irregularities would have to be young since their original amplitude could not possibly have been much larger, and diffusion along the magnetic field will quickly reduce their peak amplitude.

Although the above calculation appears to support a picture that the amplitude scintillations are produced by young, relatively intense irregularities, and this is consistent with the measured length of 6.1 km, there still seem to be problems with this interpretation of the measurements; The intensity of the irregularities, which would be a relatively

large perturbation of the background density, might be difficult to explain. However, the main difficulty seems to be that no known formation mechanism produces such short irregularities. Gradient drift type instability processes produce field-aligned structures which would be much longer than 6.1 km when they are formed, and even particle precipitation would have problems producing F region irregularities which are so short.

Because of these problems it is necessary to examine alternative ways of interpreting these measurements. One possibility is that somehow the length measurements have not allowed for some factor which is reducing the correlations and therefore giving short lengths. However, the several variations in the experimental setup which were used and the failure to detect the same irregularities if the spacing was too large seems to have confirmed the basic length measurements.

A final possible explanation is that the irregularities are actually long and are seen where gravity waves "pinch" the ionization, thereby enhancing the perturbation over a small height range. This explanation seems unlikely since irregularities are commonly seen almost simultaneously over a range of heights [MacDougall, 1981].

CONCLUSIONS

The lengths of the ionospheric irregularities which cause amplitude scintillations have been measured at 150 MHz using an array of receivers with wide spacing. The length (major-axis radius) was found to be 6.1 km. This result is much shorter than expected.

Alternative explanations for this result are:

1. The irregularities are actually long but some unaccounted for effect is making them appear short when measured on the ground. A number of different geometrical effects are involved in this experiment, but these have been taken into account.
2. The irregularities are actually long but gravity waves are "pinching" the ionization so that they are seen as short

features. It seems unlikely that this could account for many of the properties of the observations.

3. The irregularities are actually short, as shown by the measurements. This means (see discussion) that the irregularities are preferentially seen just after formation when they represent a relatively large perturbation of the background ionization. While the irregularity properties for this explanation seem consistent, there does not seem to be any known mechanism for producing such short irregularities.

REFERENCES

- Briggs, B. H., and I. A. Parkin, On the variation of radio star and satellite scintillations with zenith angle, *J. Atmos. Terr. Phys.*, 25(6), 339-366, 1963.
- Briggs, B. H., G. L. Phillips, and D. H. Shinn, The analysis of observations on spaced receivers of the fading of radio signals, *Proc. Phys. Soc. London, Ser. B*, 63, 106-121, 1950.
- MacDougall, J. W., Distributions of the irregularities which produce ionospheric scintillations, *J. Atmos. Terr. Phys.*, 43(4), 317-325, 1981.
- MacDougall, J. W., Elongation of midlatitude scintillation irregularities, *J. Atmos. Terr. Phys.*, 52(2), 151-160, 1990.
- Meek, C. E., Triangle size effect in spaced antenna wind measurements, *Radio Sci.*, 25(4), 641-648, 1990.
- Phillips, G. J., and M. Spencer, The effects of anisometric amplitude patterns in the measurements of ionospheric drifts, *Proc. Phys. Soc. London, Ser. B*, 68, 481-492, 1955.
- Vickrey, J. F., and M. C. Kelley, The effect of a conducting E layer on classical F region cross field plasma diffusion, *J. Geophys. Res.*, 87(A6), 4461-4468, 1982.
- Yeh, K.C., and C.-H. Liu, Radio wave scintillations in the ionosphere, *Proc. IEEE*, 70, 324-360, 1982.

J. W. MacDougall, Department of Electrical Engineering, University of Western Ontario, London, Ontario, Canada, N6A 5B9.

Radio wave dissipation in turbulent auroral plasma during the precipitation of energetic electrons

E. V. Mishin, L. N. Lukjanova, and S. F. Makarenko

Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, Troitsk, Russia

B. M. Atamanjuk

Fluid Mechanics Department, I.F.T.R. Polish Academy of Sciences, Warsaw

(Received February 5, 1991; revised August 16, 1991; accepted September 20, 1991.)

The results of the theoretical analysis of anomalous (collisionless) radio wave absorption in the turbulent auroral ionosphere during the intrusion of energetic electrons (i.e., in aurorae) are presented. The implications of the plasma turbulent layer (PTL) theory are used. It is shown that the dissipation of radio waves with frequencies much higher than the plasma frequency is caused by the nonlinear (combined) scattering in turbulent plasma of the PTL. In the auroral electrojet layer the principal dissipative process for the radio waves with frequencies close to the plasma frequency is $O-Z$ transformation on the field-aligned, small-scale density fluctuations. The typical dissipation decrements are estimated.

1. INTRODUCTION

In recent years, reliable experimental confirmations of the important role of plasma microturbulence in forming the fine altitudinal structure of the aurorae were obtained [Stenbaek-Nielsen and Hallinan, 1979; Dzybenko *et al.*, 1980; Mishin *et al.*, 1981, 1989; Hallinan *et al.*, 1985; Mishin and Telegin, 1989; Wahlund *et al.*, 1989]. The basic concept follows. As a result of the beam plasma instability (BPI) development, the plasma Langmuir turbulence in the E region is generated. Its intensity increases in a rather thin layer, named the plasma turbulence layer (PTL) [Mishin and Telegin, 1989]. The Langmuir waves can accelerate ionospheric electrons up to ionization energy, so the increase of the turbulence energy in the PTL results in the enhancement of the luminosity and the ionization rate. Altitude and thickness of the PTL depend on the BPI parameters. Consequently, the different altitudinal profiles of the auroral emission and ionization arise (for example, see Figure 2 of paper by Wahlund *et al.* [1989] and Figures 4 and 5 of the review by Mishin and Telegin [1989]).

For the sake of clarity, we consider the anomalous collisionless absorption of radio waves arising

from two humped electron density profiles. Of course, the existence of these effects does not directly depend on the concrete form of the $n(h)$ profile but results mainly from the development of the certain plasma instabilities. However, we believe that the consideration of the different mechanisms of absorption on the basis of this unified electron density profile illustrates the possible radio physical applications of the PTL concept to the greatest extent.

In accordance with the PTL theory [Izhovkina and Mishin, 1979; Volokitin and Mishin, 1979; Mishin and Telegin, 1989], for relatively high energy ($\epsilon_0 \geq 5$ KeV) of precipitating electrons the altitudinal profile of the turbulence can consist of two maxima (two layers) at heights of the order of 110–130 km. The upper layer (i.e., the PTL) is directly concerned with the increase of the Langmuir turbulence energy due to reduction of the nonlinear dissipation of the plasma waves excited by the precipitating electrons. The lower layer originates from the instability of an auroral electrojet flowing in the region of the auroral forms and exceeding the instability threshold [e.g., Fejer and Kelley, 1980].

Note that the advent of such a turbulence profile, as shown by Mishin and Timofeev [1981], could explain the double-peaked structure of the radio aurorae observed in the experiments of Timofeev and Miroshnikov [1982]. It is worth mentioning that

Copyright 1992 by the American Geophysical Union.

Paper number 91RS02631.
0048-6604/92/91RS-0263\$08.00

backscatter in the PTL is caused by the ion-sound type plasma density fluctuations generated as a result of nonlinear mode coupling in strong Langmuir turbulence [e.g., Sagdeev, 1979]. The alternative explanation of the effect based on the radio wave ionospheric refraction [Uspensky, 1985; Uspensky et al., 1986] is hardly applied for the 90-MHz radar observations of Timofeev and Mirosnikov [1982] mainly because of the characteristic altitudes of the peaks and the equality of their widths. But this discussion is far from the aim of our paper. The existence of a small-scale turbulence in the auroral ionosphere, as shown below, can play an important role for radio waves damping.

2. EFFECTIVE ABSORPTION OF RADIO WAVES IN THE PTL

First, let's consider anomalous absorption of the electromagnetic wave (t wave) in the upper layer, i.e., in the PTL. It is known [Tsyrovich, 1977] that during the propagation of t wave with a frequency ω_0 exceeding the plasma frequency ω_p through turbulent ($n = n_0 + \sum n_k$) plasma the effective absorption takes place as a result of t wave transformation into electrostatic l modes ($t + n_k \rightarrow l$) and because of combined (nonlinear) scattering ($t + n_k \rightarrow t$). We consider here only the second effect because the first one is negligible in the case $\omega_0 \gg \omega_p$ [Atamanjuk and Mishin, 1986].

Effective absorption of radio wave $\omega_0 \gg \omega_p$ in turbulent weakly magnetized ($\omega_p \gg \Omega_e$) plasmas is described as follows [Tsyrovich, 1977; Dawson and Oberman, 1962]:

$$\frac{d}{dt} |E_t|^2 = -\frac{\omega_p^2}{\omega_0^2} \nu_{ef}^{(t)} |E_t|^2 \quad (1)$$

$$\nu_{ef}^{(t)}(\omega_0, \mathbf{k}_0) \approx \nu + \frac{\pi \omega_p^2}{2 \omega_0} \sum_{\sigma} \int d\mathbf{k}' \frac{W_{\sigma}(\mathbf{k}')}{nT_e} \cdot \left\{ \frac{[\mathbf{k}_0 - \mathbf{k}', \mathbf{e}_t]^2}{|\mathbf{k}_0 - \mathbf{k}'|^2} \times \delta[\varepsilon_t(\omega_0, \mathbf{k}_0 - \mathbf{k}')] - \frac{c^2(\mathbf{k}_0 - \mathbf{k}')^2}{\omega_0^2} \right\} + (\mathbf{k}' \rightarrow -\mathbf{k}') \quad (2)$$

or for isotropic turbulence

$$\nu_{ef}^{(t)} \approx \nu + \pi^2 \frac{\omega_p^2}{c} \sum_{\sigma} \int_0^{2k_0} dk' k' \frac{W_{\sigma}(k')}{nT_e} \quad (3)$$

where $W_{\sigma}(k)$ is the spectral energy density of the electrostatic fluctuations (n_k); $\mathbf{E}_t = \mathbf{e}_t E_t$ is the electric field of the t wave with the frequency ω_0 and the wave vector \mathbf{k}_0 ; $n(T_e)$ is the plasma density (temperature); \mathbf{k}' is the turbulence wave number; $\varepsilon_t(\omega_0, \mathbf{k}) = 1 - (\omega_p^2/\omega_0^2)$; $\delta(x)$ is the delta function; and $\nu = \nu(T_e)$ is the ionospheric electron collision frequency.

Combined scattering takes place when $k' < 2k_0 = 2\omega_0/c$. So we can determine the lower border of the band ω_0/ω_p , where the effective absorption is essential for given distribution $W_{\sigma}(k)$. Let's consider the case of the strong turbulence [e.g., Sagdeev, 1979], which is principal for the PTL formation [Izhovkina and Mishin, 1979; Volokitin and Mishin, 1979]. At the present time the turbulence spectrum is not known in detail, and one can only suppose its main features [e.g., Shapiro and Shevchenko, 1984 and references therein]. However, as seen from (3), if the maximal wave number of the turbulence $k_{\max} < 2k_0$, the detailed shape of the spectrum is not very important [Atamanjuk and Mishin, 1986].

In the strong turbulence approach, one has two different spectral regions [e.g., Sagdeev, 1979]: (1) the "pump" region, $k \leq k_b = (\omega_p/u)$, where $u = (2\varepsilon_0/m)^{1/2}$ is the beam mean velocity; and (2) the region of the strong turbulence $k \leq k_L$, where $k_L = \lambda_D^{-1} [W_L/(3nT_e)]^{1/2}$ is the characteristic wave number of the turbulence and W_L is the turbulence energy density.

Above the PTL, i.e., for $\nu < \nu_* = (\gamma_b \omega_p m/M)^{1/2}$, the "collisionless" regime takes place where [Sagdeev, 1979]:

$$W_L \rightarrow W_L^{(<)} \approx 3nT_e \gamma_b / \omega_p \quad (4)$$

and the "pump wave" energy density

$$W_b^{(<)} \approx 3nT_e (\gamma_b / \omega_p)^{1/2} (m/M)^{1/2}$$

Here γ_b is the BPI growth rate and M/m is the ion-to-electron mass ratio. Note that the value ν_* is the inverse characteristic time of collapse of the plasma cavities in the collisionless regime of the strong turbulence [Sagdeev, 1979; Shapiro and Shevchenko, 1984].

Inside the PTL, i.e., for $\nu_* < \nu < \gamma_b$, one gets the

"collisional" regime of BPI [Mishin and Telegin, 1989], where

$$W_L \rightarrow W_L^{(>)} = 3(\nu/\omega_p)^2 n T_e M/m \quad (5)$$

$$W_b^{(>)} = 3n T_e \nu/\omega_p$$

It's seen from (4) and (5) that

$$W_L^{(>)} / W_L^{(<)} = (\nu/\nu_*)^2 > 1$$

Generally speaking, it is valid for relatively low values of the collisional frequency, i.e., $\nu < \nu_{**} \approx 0.03 \omega_p (m/M)^{1/2} \approx 10^{-4} \omega_p$ [Mishin et al., 1989]. In the opposite case, so-called "superstrong" turbulence takes place [Shapiro and Shevchenko, 1984]. Then, the following expression for w_b can be obtained [Volokitin and Mishin, 1979; Papadopoulos, 1975]:

$$W_b = (0.2 - 0.3) W_L \approx n T_e (\gamma_b/\omega_p)^2 M/m \quad (6)$$

It follows from (5) and (6) that the condition k_b (or k_L) $< 2k_0$ is fulfilled if

$$\frac{\omega_0}{\omega_p} = \frac{1}{2} \left(\frac{mc^2}{2\varepsilon_0} \right)^{1/2} \approx 2.5 (\hat{\varepsilon}_0)^{-1/2} \quad 2k_0 \geq k_b \quad (7a)$$

$$\frac{\omega_0}{\omega_p} \geq \frac{1}{2} \frac{\nu}{\omega_p} \left(\frac{Mc^2}{T_e} \right)^{1/2} \approx 10 \quad 2k_0 \geq k_L \quad (7b)$$

$$[\hat{\varepsilon}_0 = 0.1 \varepsilon_0 (\text{KeV})]$$

Therefore, though $W_L \gg W_b$, the major part of dissipation for lower frequencies $\omega_0/\omega_p < 10$ is caused by the "pump" region. The deposit of the short-scale region $k \approx k_L$ becomes important for higher frequencies $\omega_0/\omega_p > 10$. However, in this case the magnitude of the effect is reduced by a factor of $(\omega_p/\omega_0)^2$.

Below we consider the case of $\omega_0/\omega_p < 10$, when the inequality (7) is fulfilled for the "pump" region only. In this case for the sake of simplicity we suppose that

$$W_\sigma(\xi) = W_b/4\pi k_b^2 \delta(\xi - k_b)$$

Then, after substituting in (3), one obtains

$$\nu_{ef}^{(i)} = 0.1 \omega_p \frac{W_b}{n T_e} (\hat{\varepsilon}_0)^{1/2} \quad (8)$$

From (5) and (8) it follows

$$\nu_{ef}^{(i)} \geq 0.2 \hat{\varepsilon}_0^{1/2} \nu$$

and therefore the effective absorption in the PTL is unimportant for relatively weak ($\gamma_b < \nu_{**}$) beam intrusion.

Instead, in the vicinity of $\nu \approx \nu_{**}$, using (6), one obtains the effective absorption rate

$$\nu_{ef}^{(i)} = 0.1 \omega_p \frac{M}{m} (\gamma_b/\omega_p)^2$$

It yields the value of $\nu_{ef}^{(i)}$ to be of the order of ν , when

$$\gamma_b \approx [\nu_{**}(\nu)]^{1/2} \hat{\varepsilon}_0^{-1/4} \approx 10^{-4} \omega_p \approx \nu_{**}$$

Thus for $\gamma_b \approx 3\nu_{**} \approx 3 \times 10^{-4} \omega_p$ the effective absorption rate $\nu_{ef}^{(i)}$ is 10 times the collisional rate ν at the given altitude.

The absorption decrement

$$\mathcal{X}_i = \int \frac{dz}{c} \frac{\omega_p^2}{\omega_0^2} \nu_{ef}^{(i)} \approx \frac{\omega_p^2}{\omega_0^2} \frac{\nu_{ef}^{(i)} l_{\parallel}}{c}$$

depends on the scale of the turbulent region l_{\parallel} .

Letting $\omega_p = 6 \times 10^7 \text{ s}^{-1}$, $\hat{\varepsilon}_0 = 1$, $\gamma_b = (3 - 6)10^4 \text{ s}^{-1}$, $\nu_{**} \approx \nu_{**} = 6 \times 10^3 \text{ s}^{-1}$ and oblique propagation ($l_{\parallel} \approx 50 \text{ km}$), one can obtain

$$\mathcal{X}_i \leq (2 - 8), \quad \frac{\omega_0}{\omega_p} \geq 2.5$$

Therefore the attenuation of the 25-MHz radio wave in the PTL can reach 10–30 dB. For the vertical propagation ($l_{\parallel} \approx 10 \text{ km}$) we have $\mathcal{X}_i \approx 0.4 - 1.6$, corresponding to a few decibels. Thus this nonlinear scattering can yield the perceptible attenuation of the HF/VHF ($f_0 \leq 50 \text{ MHz}$) radio waves propagating through PTL in addition to the usual (collisional) absorption. This can be applied to riometric observations, oblique HF/VHF coherent scatter, and ionosondes.

3. DISSIPATION OF RADIO WAVES DUE TO $O \rightarrow Z$ TRANSFORMATION

Let's consider now the peculiar features of the energy dissipation in the lower layer. The small-scale, field-aligned irregularities with typical wave lengths of the order of one meter arising there are due to the development of, say, the Farley-Buneman instability, which is widely accepted for the

radio aurorae [e.g., *Fejer and Kelley*, 1980]. Such a structure can strongly affect propagation of radio waves due to mode coupling.

Near the ordinary (o) wave reflection point the transformation into very slow quasi-longitudinal waves (so-called z mode) can take place besides usual negligible $o \rightarrow o$ scattering. Earlier this approach was used for the explanation of the anomalous dissipation of the ordinary waves in the active experiments with HF heating [*Mityakov et al.*, 1975] and in the middle-latitude ionosphere F region [*Denisenko et al.*, 1987a]. The existence of this mechanism in the middle latitudes is confirmed by the experimental data [*Eccles and King*, 1970; *Denisenko et al.*, 1987b].

We use o - z transformation for describing the interaction between radio waves and natural, small-scale irregularities in the auroral electrojet layer. It is essential that the radio waves are considered as the "test" waves interacting with given spectrum of the irregularities. In other words, the amplitudes of the radio waves are below the threshold of the parametric instabilities.

In this case the polarization vector of both the incident and scattered waves must have only a small angle with the plane orthogonal to the geomagnetic field, which corresponds to the quasi-parallel propagation for o wave and quasi-orthogonal for z mode. Then, we have the following transport equation for spectral density of the energy of the o wave in the vicinity of the reflection point [*Ryzhov*, 1972; *Mityakov et al.*, 1975]:

$$\frac{d}{dz} (v_{gr} \omega_0) |E_0|^2 = -\gamma \omega_0 |E_0|^2 \quad (9)$$

$$\gamma = \frac{8\pi^2 \omega_p^4 \Omega_e^2 \sin^2 \psi}{\omega_0 (\omega_p^2 + \Omega_e^2 - \omega_0^2)^2} \cdot \int_{-\infty}^{\infty} (k_{\perp}^2 + q_{\parallel}^2) \delta N_q^2(q_{\perp}^{(0)}, q_{\parallel}) dq_{\parallel} \quad (10)$$

where $\delta N_q^2 = \Delta N_q^2 / N^2$ is the spectrum of the relative plasma density fluctuations; $k_0 = \omega_0 / c$; q_{\perp} , q_{\parallel} are correspondingly orthogonal and parallel (with respect to B_0) components of the wave number of the irregularity; $v_{gr} = \partial \omega_0 / \partial k_0$, ψ is the angle between the electric field of the fallen wave E_0 and B_0 .

$$q_{\perp}^{(0)} = k_{\perp} \approx \left[\frac{k_0^2 \Omega_e^2 + q_{\parallel}^2 (\omega_0^2 - \omega_p^2)}{\omega_0^2 - \omega_p^2 - \Omega_e^2} \right]^{1/2} \quad (11)$$

It is implied that $1 - \Omega_e^2 / \omega_0^2 < \omega_p^2 / \omega_0^2 < 1$.

Consider the simple example of the power law spectrum of δN_q^2 on the transverse wave numbers $\delta N_q^2 = \delta N_q^2(\delta N_q^2)$; $\delta N_q^2 = A q_{\perp}^{-2\beta}$; ($q_1 \leq q_{\perp} \leq q_2$),

where

$$A = (\beta - 1) [q_1^{2(1-\beta)} - q_2^{2(1-\beta)}]^{-1} \frac{\delta N_0^2}{\int_{-\infty}^{\infty} \delta N_q^2 dq_{\parallel}}$$

Then, for the linear model of the layer $\omega_p^2 / \omega_0^2 = 1 - z/L$, reasonable in the E region for $L \leq 10$ km, one can estimate the integral o wave relaxation due to the transformation into z mode:

$$\mathcal{X}_{o \rightarrow z} = \int_0^{L/\Omega_e} \frac{\gamma dz}{v_{gr}} \quad (12)$$

Supposing the angle $(k_0 B_0) \ll 1$ (or $\psi \approx 90^\circ$), $q_1 \ll q_2$, $\Omega_e \ll \omega_0$, one obtains

$$\mathcal{X}_{o \rightarrow z} \approx 8\pi \left(\frac{q_1}{k_0} \right)^{2(\beta-1)} \delta N_0^2 \left(\frac{\omega_0}{\Omega_e} \right)^{1/2} k_0 L \quad (13)$$

In (12), letting $q_1 \approx 5 \times 10^{-2} \text{ cm}^{-1}$, $\beta \approx 1.5$, $\delta N_0^2 \approx 10^{-4}$, $\Omega_e \approx 8 \times 10^6 \text{ s}^{-1}$, $\omega_0 \approx \omega_p \approx 3 \times 10^7 \text{ s}^{-1}$, and $L \approx 10$ km, one has

$$\mathcal{X}_{o \rightarrow z} \approx 10$$

Note that the value of $\delta N_0 \approx 10^{-2}$ is, in fact, close to a minimal value in the perturbed ionosphere. Indeed, to explain the observations of the radio aurorae and of the auroral plasma heating, the values of δN_0 at least a few times higher are needed [e.g., *Fejer and Providakes*, 1987]. Therefore o - z transformation in the electrojet layer can be rather significant for damping of the HF radio waves propagating almost along the geomagnetic field lines.

4. CONCLUSIONS

We have considered two mechanisms of the collisionless absorption of the HF/VHF radio waves propagating in the auroral ionosphere during the precipitation of an intensive beam of energetic electrons characteristic for the active auroral forms. For the radio waves with frequencies much higher than the plasma frequency the dominant process is the nonlinear (combined) scattering on the Lang-

muir waves in the layer of the plasma turbulence excited by the precipitating electrons.

The independent process is the transformation of the radio waves near the reflection point into z mode on the small-scale, field-aligned irregularities typical for the radio aurorae. It takes place for the radio waves propagating almost vertically in the auroral electrojet layer where the irregularities are generated. It is evident that in the latter case the frequency of the radio wave does not exceed the plasma frequency in the E region maximum, i.e., $f_0 \leq 10$ MHz.

It is worth mentioning that when the radio wave frequency is less than the double-plasma frequency in the maximum of PTL, the anomalous damping in the PTL increases as a result of the transformation into Langmuir waves on the plasma turbulence. We did not include this effect in the paper. We would like to emphasize that the localization of the PTL is inspected by the BPI parameters. The PTL can coincide, for instance, with the layer of an auroral electrojet, so that both mechanisms would "work" simultaneously in the same place. One can expect the anomalous dissipation of the radio waves to be much more effective in this case.

REFERENCES

- Atamanjuk, B. M., and E. V. Mishin. On mechanism of radio wave absorption in ionosphere during electron beam injection (in Russian). *Geomagn. Aeron.*, 26(4), 665-667, 1986.
- Dawson, J., and C. Oberman. High frequency conductivity and the emission and absorption coefficients of a fully ionized plasma. *Phys. Fluids*, 5, 517-524, 1962.
- Denisenko, P. F., V. I. Vodolazkin, Yu. N. Faer, and L. N. Boltyukhova. Radio waves anomalous absorption and effective collisions frequency in the ionosphere F -region (in Russian). *Geomagn. Aeron.*, 27(3), 504-506, 1987a.
- Denisenko, P. F., N. A. Zabolotn, S. A. Pulnits, V. V. Selegei. Transformation of an ordinary wave into extraordinary wave on the topside ionosphere sounding data (in Russian). *Geomagn. Aeron.*, 27(4), 544-549, 1987b.
- Dzyubenko, N. I., V. N. Ivchenko, G. P. Milinevsky, and E. V. Mishin. Influence of plasma collective effects on the auroral rays structure (in Russian). *Zh. Eksp. Teor. Fiz.*, 31(11), 643-646, 1980.
- Eccles, D., and J. W. King. Ionospheric probing using vertical incidence sounding techniques. *J. Atmos. Terr. Phys.*, 32, 517-538, 1970.
- Fejer, B. G., and M. C. Kelley. Ionospheric irregularities. *Rev. Geophys.*, 18(2), 401-454, 1980.
- Fejer, B. G., and J. F. Providakes. High latitude E -region irregularities: New Results. *Phys. Scr.*, 18, 167-178, 1987.
- Hallinan, T. J., H. C. Stenbaek-Nielsen, and C. S. Deehr. Enhanced aurora. *J. Geophys. Res.*, 90(A9), 8461-8475, 1985.
- Izhovkina, N. I., and E. V. Mishin. On possibility of beam plasma discharge ignition during intrusion of auroral electrons in the ionosphere. *Geomagn. Aeron.*, Engl. Transl., 19(3), 398-399, 1979.
- Mishin, E. V., and V. A. Telegin. Effects of plasma turbulence in Aurorae. *Geomagn. Aeron.*, 29(1), 1-13, 1989.
- Mishin, E. V., and E. E. Timofeev. On possible mechanism of double-peaked radioaurora formation (in Russian). *Geomagn. Aeron.*, 21(1), 201-202, 1981.
- Mishin, E. V., V. N. Ivchenko, and G. P. Milinevsky. Fine structure of artificial auroral rays. *Adv. Space Res.*, 1(2), 163-165, 1981.
- Mishin, E. V., Yu. Ya. Ruzhin, and V. A. Telegin. *Interaction of Electron Fluxes With the Ionospheric plasma* (in Russian), 264 pp., Hydrometeoizdat, Leningrad, 1989.
- Mityakov, N. A., V. O. Rapoport, and V. Yu. Trakhtengerts. Scattering of an ordinary near the reflection point by small-scale irregularities (in Russian). *Izv. Vyssh. Uchebn. Zaved. Radiofiz.*, 18(9), 1273-1275, 1975.
- Papadopoulos, K. Nonlinear stabilization of beam plasma interactions by parametric effects. *Phys. Fluids*, 18(12), 1769-1777, 1975.
- Ryzhov, Yu. A. Electromagnetic field energy absorption in the chaotically inhomogeneous plasma (in Russian). *Zh. Eksp. Teor. Fiz.*, 62(3), 924-931, 1972.
- Sagdeev, R. Z. The 1976 Oppenheimer lectures: Critical problems in plasma astrophysics. 1. Turbulence and nonlinear waves, 2. Singular layers and reconnection. *Rev. Mod. Phys.*, 51(1), 1-20, 1979.
- Shapiro, V. D., and V. I. Shevchenko. Strong turbulence of plasma oscillations. in *Foundations of Plasma Physics* (in Russian), vol. 2, edited by A. Galeev and R. Sudan, pp. 119-174. Energoatomizdat, Moscow, 1984.
- Stenbaek-Nielsen, H. C., and T. J. Hallinan. Pulsating aurora: Evidence for non-collisional thermalisation of precipitating electrons. *J. Geophys. Res.*, 84(7), 3257-3272, 1979.
- Timofeev, E. E., and Yu. Miroshnikov. Altitude characteristics of radar aurora as seen by 90 MHz double-altitude radar system operated at Karmaselga, Karelia. *J. Geophys.*, 51(1), 44-54, 1982.
- Tsytovich, V. N. *Theory of Turbulent Plasma*, 422 pp., Plenum, New York, 1977.
- Uspensky, M. V. On the altitudinal profile of auroral radar backscatter. *Radio Sci.*, 20(4), 735-739, 1985.
- Uspensky, M. V., M. K. Vallinkosky, and T. Turunen. On possibility of the formation of the double-spike altitudinal auroral backscatter profile (in Russian). *Geomagn. Aeron.*, 26(4), 595-599, 1986.
- Volokitin, A. S., and E. V. Mishin. On electron beam relaxation in the plasma with rare collisions (in Russian). *Fiz. Plazmy Moscow*, 5(5), 1166-1169, 1979.
- Wahlund, J. E., H. J. Opgenoorth, and P. Rothwell. Observation of thin auroral ionization layers by EISCAT in connection with pulsating aurora. *J. Geophys. Res.*, 94(A12), 17,223-17,233, 1989.
- L. N. Lukjanova, S. F. Makarenko, and E. V. Mishin. Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, 14209, Troitsk, Moscow, Russia.

NICARE I HF propagation experiment results and interpretation

P. Argo, T. Joseph Fitzgerald, and Robert Carlos

Los Alamos National Laboratory, University of California, Los Alamos, New Mexico

(Received February 12, 1991; revised October 7, 1991; accepted October 7, 1991.)

The NICARE I chemical release created a large electron density depletion in the nighttime *F* region of the ionosphere at an altitude of approximately 300 km which persisted for at least 30 min. A three-dimensional ray-tracing code (TRACKER) predicted that the lenslike refractive perturbation would create new off-axis modes in long-distance HF radio paths which reflected in the ionosphere near the depletion. Channel measurements during this experiment bear this out. The multipath created by these new modes induced an increased fading rate on the received signals. Analysis of the time history of the direction of arrival of the new modes indicates that after the initial expansion of the release the western edge continued to move westward while the eastward edge remained stable in location. This behavior is consistent with model predictions of the behavior of the depletion.

INTRODUCTION

Since the 1960s ionospheric modification experiments in which electron density depletions were generated by the high altitude release of chemicals with high electron attachment have been conducted. The depletion studies have led to various models of the attachment chemistries (see, for example, *Mendillo and Forbes* [1982] and *Bernhardt* [1987]). These experiments have had profound effects on nearby ionosondes [*Booker*, 1961] but have not shown that the depletions affect long path oblique HF propagation. See *Mendillo* [1988] for a comprehensive review of ionospheric depletion theory and experiments.

We have developed a simulation model, which uses a three-dimensional ray-tracing program (TRACKER, developed at Livermore and Los Alamos National Laboratories and described briefly in Appendix A) to calculate the path of radio energy through the ionosphere perturbed by a simplified chemistry model derived from the work of *Bernhardt* [1987]. Our computer experiments indicated that under optimum conditions we would be able to detect effects of the depletion on the propagation channel, in particular new modes would be generated and existing modes would undergo modified Doppler shifts and group path changes. Los Alamos National Laboratory participated in the NICARE I release experiment with the express purpose of

measuring the effects of the electron density hole on oblique HF propagation. We fielded a suite of experiments designed to quantify the HF propagation channel characteristics (e.g., frequency coherence bandwidth, spatial correlation distance). Previous experimenters gave measured HF propagation channels through quiet or naturally disturbed ionospheres [*Basler et al.*, 1988; *David et al.*, 1969; *Taheri and Steinberg*, 1979], using spaced antennas and frequencies to quantify the channel via frequency coherence bandwidths, spectral correlation distances, and fading rates. With the availability of three-dimensional ray-tracing codes [*Jones*, 1968], many computer experiments on the effects of gradients and structures have been conducted [e.g., *Helms and Thompson*, 1973]. This paper will discuss measurements of the relevant channel effects (angle of arrival changes, phase path changes (Doppler), group path changes, fading rates, and frequency coherence bandwidths). At the end we will compare the modeling of the NICARE propagation channel to the measurements.

In this paper we describe the impact of the NICARE I chemical depletion upon long-distance HF paths which reflected in the ionosphere near the depletion. The depletion was produced by the release of CF_3Br from a rocket launched at Wallops Island, Virginia, at 0158 UT on October 23, 1989. The release occurred at 37.0038 N, 73.6676 W and an altitude of 310 km in the nighttime *F* layer. We will concentrate on the experiments fielded by Los Alamos which consisted of multiple-frequency CW transmissions over two paths: a 1740-km path be-

Copyright 1992 by the American Geophysical Union.

Paper number 91RS02916.
0048-6604/92/91RS-02916\$08.00

NICARE

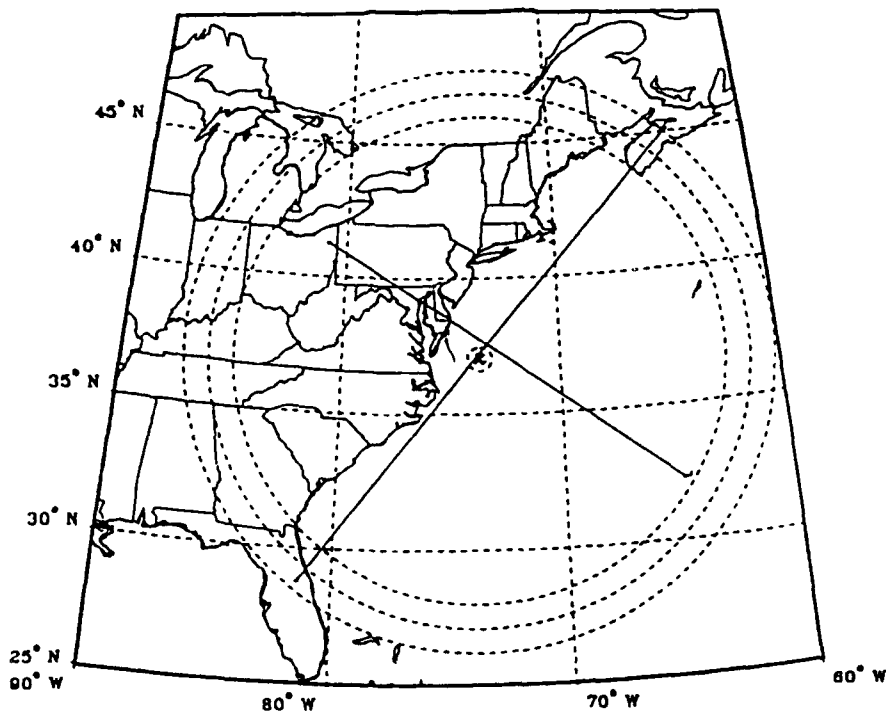


Fig. 1. Map indicating the location of the CF_3Br release in relation to the great circle paths between Bermuda and Ohio, and between Florida and Nova Scotia.

tween Bermuda (32.28 N, 64.85 W) and Warren, Ohio (41.27 N, 80.92 W), and a 2380 km path between Sandford, Florida (28.78 N, 81.44 W), and Stanley, Nova Scotia (45.10 N, 63.92 W) which are shown on the map in Figure 1. Significant effects were observed on both paths, especially between Florida and Nova Scotia for which the depletion occurred near the great circle connecting the stations. We also will briefly describe the results of other experiments which add supplementary information. These include the NRL channel probe deployed between Bermuda and Ohio, the Wallops Island ionogram, and the Millstone Hill incoherent scatter radar.

EXPERIMENTAL SETUP

The NASA supported NICARE (Nickel Carbonyl Release Experiment) I chemical depletion

experiment consisted of two 310 km altitude chemical releases from a single rocket launched at Wallops Island, Virginia, on the night of October 22 (local), 1989. The chemicals released were 30 kg each of nickel carbonyl and trifluoromethyl bromide (halon). The depletion chemistry has been discussed by *Bernhardt* [1987].

Los Alamos fielded long path bistatic HF propagation links from Bermuda to Ohio, and Florida to Nova Scotia (Figure 1). These crossed paths were selected to place the midpoint of the propagation paths near the pair of ionospheric depletions. Each transmitter site broadcasted four 100-W CW tones with frequencies stabilized by a rubidium oscillator. In general, the frequencies were spaced at 10, 30, and 70 kHz above the base frequency. The tones were fed in pairs (0–10 kHz, 30–70 kHz) into two wideband antennas suspended from one mast. One antenna was an end-fired delta aligned so that the

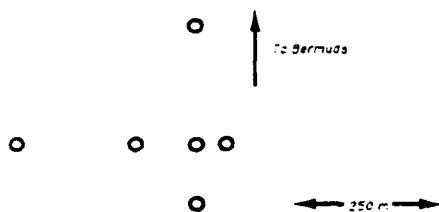


Fig. 2a. Spatial locations of the Warren, Ohio, array elements.

plane of the antenna was also the plane of the nominal ray path. The other transmit antenna was a tilted delta with the plane of the antenna perpendicular to the other antenna and containing the nominal ray to the receiver station. Thus the transmit tones were linearly polarized, but with opposite polarizations for the different antennas. Each receiver site had an array (Figures 2a and 2b) of magnetic loop antennas with a diameter of 1 m, each feeding a separate RACAL RA6790/GM receiver. In addition, one antenna element was used to receive all four frequencies. The receivers were operated at a 10-Hz beat frequency offset and an audio bandwidth of less than 100 Hz. The receiver audio outputs were each sampled at 100 Hz with a 12-bit analog to digital converter; the digitized data was stored on magnetic disk. These spaced-antenna/frequency measurements allow us to monitor the changes in the propagation channel characteristics—in particular the frequency coherence bandwidth and the spatial coherence lengths.

In addition to the Los Alamos HF instrumentation there were other experiments used to add supplementary information. The Naval Research Laboratory broadband channel probe [Wagner and Goldstein, 1983] was deployed on the Bermuda to Ohio path in order to provide group path information. The Digisonde 256 which is routinely operated at Wallops Island took ionograms on a five minute schedule; unfortunately, it did not include data on angle-of-arrival information. The Millstone Hill incoherent scatter radar was scheduled to measure the size, depth, and location of the pair of depletions. There were also in situ plasma diagnostics on a daughter payload that passed through the releases a few seconds (and kilometers) behind the release payload. Los Alamos fielded two optical sites (Wallops Island and Duck, North Carolina) to provide independent measurements on the depletion sizes and locations.

The first cannister released nickel carbonyl slowly rather than in the burst as planned, and so did not produce the expected depletion. The second release appeared to go as planned. Although in situ measurements located within a few kilometers of the two releases indicated strong local depletion effects (from both releases), neither the incoherent scatter radar nor the two optical stations showed any clear sign of a depleted region for either chemical release. The optical signature for the second release was masked by the bright glow of the continually released nickel carbonyl. The incoherent scatter radar tracked the rocket beyond the second release and after returning to the release point remained there long in an integration mode. Meanwhile, the depletion had moved westward. The Wallops Island ionosonde showed a new reflection following the halon release, indicating that major ionospheric changes had occurred.

WALLOPS ISLAND VERTICAL INCIDENCE SOUNDER

A vertical incidence ionosonde was run at Wallops Island (37°N, 74°W) at a range of 200 km from the release. These ionograms which were taken at 5-min intervals show the development of a disturbance, seen as a secondary trace on the series of ionograms (see Figure 3, for example). The first ionogram showing such a trace was at 0209 UT, approximately 5 min after the release. The delay of the secondary trace compared to the vertical trace indicates a reflection at an oblique angle from a tilted ionosphere. We have inverted from each ionogram both the main trace and the tilted trace. The tilted trace has initially been treated as a vertical trace by the inversion process to get true ranges from the ionosonde, and then is directed back (off vertical) toward the east to estimate the density perturbations. Unfortunately, the Di-

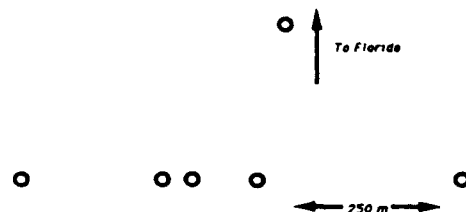


Fig. 2b. Spatial locations of the Stanley, Nova Scotia, array elements.

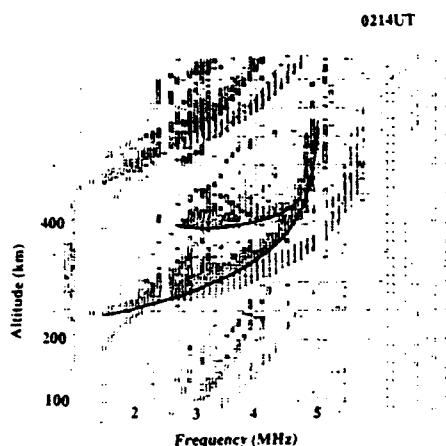


Fig. 3. Ionogram obtained at Wallops Island at 0214 UT showing a delayed secondary trace indicating an off-vertical reflection induced by the electron depletion.

gisonde 256 run at Wallops Island was not run in a mode that measures the angle of arrival of the echoes. This would have been useful in tracking the depletion with time.

We used the vertical trace to define a horizontally "layered" ionosphere for each time and then applied the "tilted" secondary trace at the appropriate angle to get density contours. Because neither the optical measurements nor the incoherent scatter radar at Millstone Hill observed a depletion, and the ionosonde could not determine directionality, we are left with few clues as to the movement of the depletion following the release. Using the bistatic CW antenna array as a direction finding tool, we ascertained that the depletion appeared to drift westward with a velocity of ~ 50 m/s. If we assume the depletion drifted toward the west (and Wallops Island) at a velocity of 50 m/s, then the sequence of reconstruction of the plasma profiles shown in Figure 4 result. These figures show the depletion growing over its first 40 min of life. From the ionosonde data we can say that a depletion in fact did occur, that it developed and enlarged over the first 40 min, and that it appeared to drift westward against the eastward neutral wind normally present in the premidnight ionosphere.

We have used TRACKER to model the effects of the depletion on a vertical incidence ionosonde at Wallops Island. The first data we have after the release was at 5 min (300 s), and the secondary trace on the ionogram showed that a frequency slightly

greater than 4 MHz was affected by the depletion. By using a model ionosphere updated to match the vertical ionogram we ray traced into depletions of varying release amounts, looking for the situation where 4-MHz rays just penetrated the depletion. This modeling indicates that the secondary trace on the ionogram is created by a nonvertical echo, one that has traveled eastward to the depletion, been reflected by the perturbed plasma contours, and returned to the ionosonde. These signals are likely to be strongly defocused, and the received signal strength will be many tens of decibels down. In fact, the amplitude coded ionogram shows this to be the case.

Modeling of this kind has indicated that the release of CF_3Br can be simulated by assuming the 30 kg of material was released at a temperature near 300°K. A description of the chemistry models is included in Appendix B.

DOPPLER SPECTRA

The CW transmissions consisted of pure HF tones generated by a frequency synthesizer which was stabilized by a rubidium clock. Because of time variations in the index of refraction of the ionosphere the phase of the receive signal will vary resulting in a small Doppler shift of the CW transmission which depends on the exact ray path. Negative Doppler shifts correspond to increasing phase paths. Thus multipath will appear as multiple peaks in the power spectra with different frequency shifts for each component. The time variation of the centroid of the peak reveals the temporal evolution of the ionosphere along that particular ray. Moreover, modes with a weak signal strength may be apparent because they are Doppler shifted away from the more dominant modes. Our method of calculating the evolution of the Doppler power spectra is based on the short-term Fourier transform. One divides the total time series of data into smaller subintervals (10–40 s) over which the spectra appear stable and calculates the spectrum during that subinterval. The length of the subinterval is adjusted to give adequate frequency resolution without smearing produced by changes during the subinterval. The subwindow is then advanced some fraction of its length and a new power spectrum calculated. The results may be visualized in a three dimensional plot which shows power at a given frequency and time as a grey level.

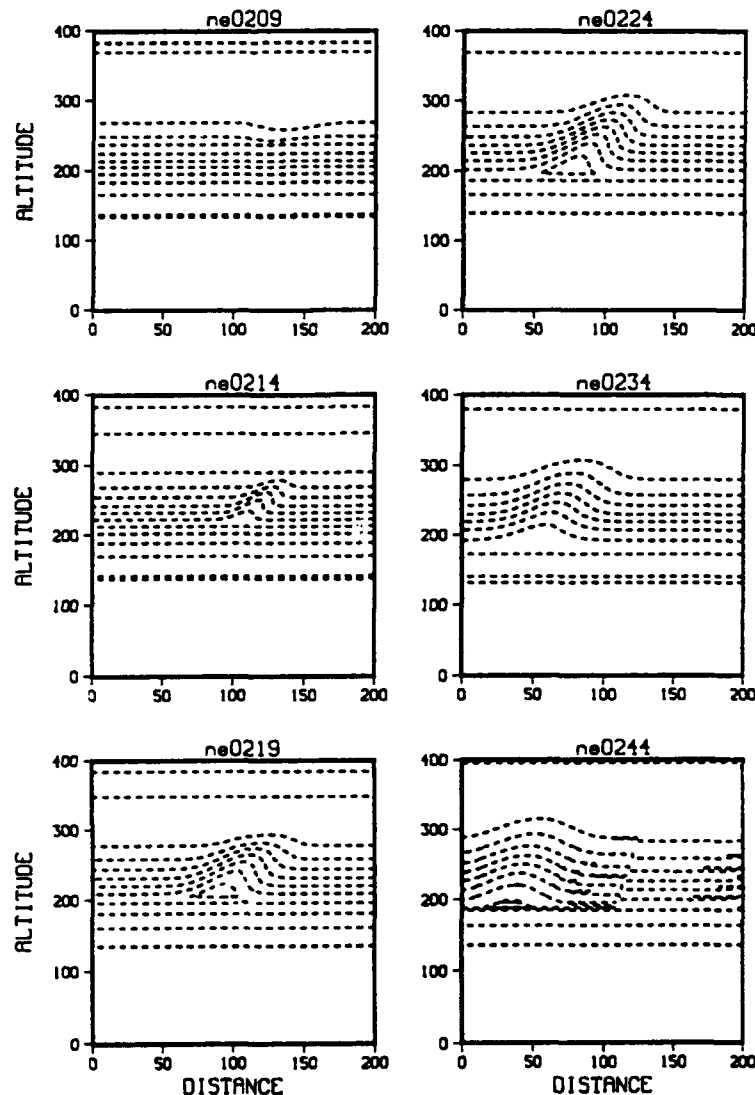


Fig. 4. Electron density contours reconstructed assuming that the depletion drifted toward the ionosonde at 50 m/s. Measurement times are 0209, 0214, 0219, 0224, 0234, and 0244 UT. Contours start at 0 MHz and are spaced by 0.4 MHz.

Doppler spectra on Florida-Nova Scotia path

The data set of major interest to the impact of the NICARE releases on the Florida to Nova Scotia path was obtained beginning at 0143:02 UT and extended to 0232:00 UT, by which time the 13.60-MHz transmission was no longer reflecting from the ionosphere for the path. The second release occurred at 0204:13. A disturbance corresponding to a weak mode with a relative high negative Doppler

was observed almost immediately after the second release. This is illustrated in Figure 5 which displays the behavior of the Doppler spectrum versus time after chemical release (0204:13 UT). The mode associated with the second release decreases in Doppler with time; it appears to cross the Doppler track of the preexisting modes at about 227 s after the release and stabilizes at a Doppler of about 0.1 Hz. The spectral peak for this path appears to be

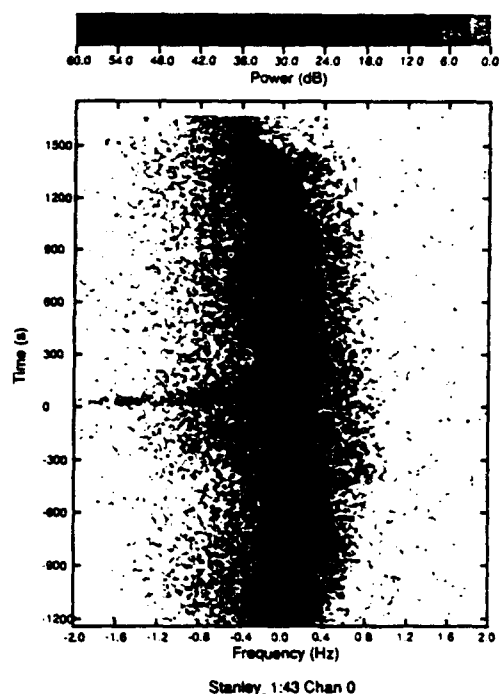


Fig. 5. Time history of power spectrum of CW transmission received at Stanley, Nova Scotia. The vertical scale represents time in seconds relative to chemical release (0204:13 UT), while the horizontal scale represents the modulation frequency in hertz. The power is encoded according to the scale at the top. A disturbance induced by the second release (0 s) appears as a new peak at negative frequency.

made up of at least two paths with slightly differing Doppler shifts. After 1000 s the spectra becomes more complicated as the path approaches penetration. Thus the behavior of the spectra in Figure 5 indicate that the second release produced new and relatively stable ray paths for at least a period of 20 min. We note that there is a significant increase in signal strength after 200 s on the mode at about -0.2 -Hz Doppler; this is also observed in the total power for all modes. There is a clear transition in the width of the spectra before and after the release; the width before the release is approximately 0.1 Hz while afterward it is 0.3 Hz. We would expect to observe this effect as an increase in the fading rate after the release.

During the period before the launch two peaks in the Doppler spectra with a separation of about 0.1 Hz indicate the presence of two modes which we ascribe to O and X rays (see below). We note that the transmission frequency was too high to support two-hop propagation; it is also unlikely that (two-

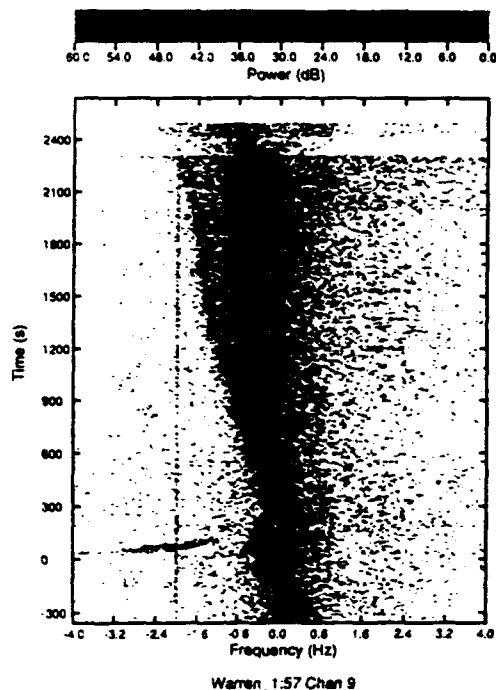


Fig. 6. Time history of power spectrum of CW transmission received at Warren, Ohio. A disturbance associated with the second release appears as highly chirped peaks at negative Doppler between 30 and 120 s.

hop) sporadic *E* propagation would be possible. No sporadic *E* was observed at this time period on the ionograms recorded at Wallops Island. The broad noise in the spectra at the 10-dB level surrounding the peaks is highly correlated between receivers. No disturbance associated with the first NICARE release is apparent in the Doppler spectra.

Doppler spectra on Bermuda to Ohio path

A data set beginning 1 s before launch (0157:59 UT) was obtained on the Bermuda to Ohio path at a transmission frequency of 10.1 MHz; it extended to 0246 UT by which time electron density at the reflection point had decreased enough that the transmission was no longer reflected. We have examined the Doppler spectra for this data set in the same manner as described above. The results are shown in Figure 6. The spectra at launch time show only a single peak near zero Doppler; data provided by the NRL Channel Probe on this path indicate that the peak is associated with the one-hop low ray. Both O and X modes appear on the delay data provided by the Channel Probe but because of

accidental polarization matching which will be described below the X mode dominates in the CW data.

Two highly chirped but weak peaks at negative Doppler beginning about 40 s appear to associated with the second release. They last for about 100 s at a power level of 10 to 20 dB; they start with a Doppler shift (magnitude) of greater than 3 Hz and decreasing to about 1 Hz. The unusual behavior of these modes leaves little doubt that they were associated with the release. The peak at about -0.6 Hz which gradually becomes well defined after 800 s and persists until the junction at 2400 s is the normal high ray. This interpretation is verified by Channel Probe and angle-of-arrival data. The low and high (X) modes merge at the junction frequency focussing event at 2700 s. A weak mode persisted after the junction. The junction of the weak O mode occurred at about 2400 s. The interpretation of the weak peak at about -0.6 Hz between 0 and 200 s is difficult; the timing seems to be associated with the release but the Doppler shift corresponds to the high ray noted before. There is also another peak at negative Doppler that begins at about 800 s near -0.6 Hz and moves out to increasingly negative Doppler before fading out after 2200 s. This peak may correspond to a weak low ray path identified in the Channel Probe data. The peak near -2 Hz that persists through the data set is likely to be caused by an interfering transmitter. In conclusion, the only clearly defined disturbance produced by the NICARE releases on the Bermuda to Ohio path was the weak peaks at negative Doppler that last for about 100 s after the second release.

ANGLE OF ARRIVAL MEASUREMENTS

The elements of each receiver station which shared a common frequency were phase coherent and therefore could be used as a synthesized radio telescope to determine angle of arrival. The Warren array consisted of seven elements in the configuration shown in Figure 2a. The maximum aperture transverse to the nominal path was 750 m; the maximum aperture parallel to the path was 300 m. Given that the elevation angle of the nominal ray was about 20° , the effective vertical aperture was approximately 100 m. Thus the resolution in azimuth was better than that for elevation angle. The Stanley array (Figure 2b) consisted of six elements with a maximum aperture of 744 m in the transverse

direction and 365 m in the longitudinal direction. Since the nominal elevation angle at Stanley was about 10° the effective vertical aperture was approximately 60 m. The difference in elevation angles at the two stations was a result of the longer range on the Stanley path and the asymmetric effect of ionospheric tilts. The smaller vertical aperture at Stanley was somewhat compensated by the use of about a 30% shorter wavelength. Since there was only one element off the transverse axis, the angle of arrival estimates are more subject to aliasing (different elevation angles giving same phase) than those of the Warren array. At Stanley, an independent subarray of two elements with the maximum transverse separation was formed at a frequency 80 kHz above that used for the full array of six elements.

Our method of calculating the angle of arrival is based on the delay-and-sum beam former algorithm using Doppler filtering to separate the multipath components. Suppose that the vector \mathbf{r} with components (x, y, z) represents the location of an antenna; we assume a coordinate system in which x and z are in the horizontal plane containing the receive antennas with z directed toward the transmitter and x transverse to the path. Then y points into the ground. Relative to an antenna at the origin, the phase of a plane wave with wave vector \mathbf{k} at a point \mathbf{r} is $\exp(-i\mathbf{k} \cdot \mathbf{r})$ for the convention we use for our data. Define η as the zenith angle of the ray and ξ as the azimuth angle of the ray relative to the nominal direction to the transmitter. Then the wave vector \mathbf{k} has the components $k(-\sin \eta \sin \xi, \cos \eta, -\sin \eta \cos \xi)$, where k is the free space wave number. If F_i is the complex amplitude at antenna i , then the output of the beam form, $f(\sin \xi, \sin \eta)$, is

$$f(\sin \xi, \sin \eta) = \sum_i F_i \exp(+i\mathbf{k} \cdot \mathbf{r}) \quad (1)$$

for an incident plane wave. We assume that the antenna radiation pattern is much broader than the expected variation in the angle of arrival. Letting $\langle F_i F_j^* \rangle$ represent the coherence of antennas i and j , the direction of arrival is obtained by searching for a maximum in the brightness

$$|f(\sin \xi, \sin \eta)|^2 = \sum_i \sum_j \langle F_i F_j^* \rangle \cdot \exp\{-ik \sin \eta [\sin \xi (x_i - x_j) + \cos \xi (z_i - z_j)]\} \quad (2)$$

which reduces to

$$|f(\sin \xi, \sin \eta)|^2 = n + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \operatorname{Re} \{ \langle F_i F_j^* \rangle \cdot \exp [-ik \sin \eta (\sin \xi (x_i - x_j) + \cos \xi (z_i - z_j))] \} \quad (3)$$

where n is the number of antennas. We calculate the coherence in the Doppler frequency domain using the Fourier transform of the time series in the interval of interest; the mode of interest may be selected by frequency filtering. The relative phases of the different channels are preserved in the frequency domain.

Angle of arrival on Florida-Nova Scotia path

Angle of arrival measurements obtained from the receiver array in Nova Scotia clearly indicate the production of azimuthal multipath following the second release. The signals from the various receivers have unknown phase shifts due to circuitry and cable length differences; we therefore cannot determine the absolute direction of arrival. However, the phase shifts remain constant for the duration of a data set so that we can determine angle of arrival of differences. To do this, we select a time period during which the spectra are quiet and assume that the elevation angle is that given by ray tracing in a model ionosphere corrected with the Wallops Island ionograms and the azimuth is broadside to the transverse arm of the array. From the measured phase differences during this time period we can find the relative phase changes needed to align the brightness, $|f(\sin \xi, \sin \eta)|^2$, with the nominal direction of arrival. We then correct the whole data set using these phase changes; our calculated directions of arrival will then be relative to the nominal ray. For the Florida-Nova Scotia path we assume that the elevation angle of the nominal ray is 11° . Ray tracing indicates that the nominal ray would arrive from slightly west of the great circle path due to ionospheric tilts. We selected a time period from 270 to 170 s before release, when the spectra showed one major peak, to determine the phase correction.

We cannot obtain an accurate direction of arrival for the weak disturbance occurring immediately after the second release. However, once the power in the negative Doppler peak increases, we can distinguish a direction distinctly to the east of the nominal path (1.6°) and at about the same elevation

angle. This direction was obtained between 20 and 50 s after release at Doppler shifts between -0.8 and -0.6 Hz. The peak near 0 Hz showed no significant change during that time. The direction of arrival of the new mode remained displaced to the east and at somewhat lower elevation angle compared to the nominal direction. Thus the second release immediately formed a weak, rapidly varying path: the negative Doppler shift indicates that the phase path was increasing. This new mode gradually decreased in Doppler shift while increasing in strength to form a relatively stable path arriving from the east of the nominal path. During approximately the first 175 s following the release the nominal path remained undisturbed; this indicates that the release was in between the two ray paths and had a diameter of less than 40 km up to at least 175 s after release.

During the period from 170 to 270 s the Doppler shifts of modes overlap so that filtering is not possible. The most likely interpretation of Figure 5 is that the modes cross during this time period; that interpretation would be consistent with the eastward displacement of the mode near 0-Hz Doppler at times between 370 and 420 s. The mode at about -0.1 Hz Doppler is displaced to the west 1° during that time period; 100 s earlier the same mode was displaced a somewhat lesser amount to the west. Apparently, by 370 s the new mode had stabilized in phase path so that the Doppler had decreased to almost zero but the direction of arrival remained displaced to the east. Sometime after 170 s the original mode started being displaced to the west; the negative Doppler indicates a growing phase path consistent with the displacement. Indeed, at 750 s there was a third mode at displaced toward the east with a somewhat higher elevation angle than the other.

Between 370 and 1100 s the spectra are highly structured but consistently show modes displaced to the east and a mode that moves increasingly to the west with time. Apparently, the eastward side of the depletion remained relatively stable while the westward side after reaching the nominal path at about 170 s continued to expand to the west for the duration of the data set. The increasing westward displacement of the most negative Doppler mode was at approximately 40 m/s transverse to the path. The most positive Doppler mode at the same time remained displaced to the east with a relatively low elevation angle.

Angle of arrival on Bermuda to Ohio path

We corrected the phases of the various Warren receivers in the same manner as described for the Stanley group assuming the nominal low ray mode had an elevation angle of 19° ; we based the phase correction on the data in the main Doppler peak at 700 s after release. The most obvious change in the spectra shown in Figure 6, the development of a strong mode at about -0.6 Hz Doppler, is distinctive in the angle of arrival results. The Doppler peak at -0.2 Hz is close to the nominal path while the peak at -0.6 Hz is distinctly above and to the south of the nominal path; these results lend credence to the identification of the peak at -0.2 Hz as the low ray and that at -0.6 Hz as the high ray. Up to the junction at about 2300 s the angle of arrival of the mode with the smaller Doppler shift (low ray) gradually approaches that of the high ray; that is, the azimuth of low ray is increasingly deflected from the nominal direction. It may be that azimuthal deflection arises from ambient ionospheric tilts; however, the deflection is in the direction of the NICARE depletion.

The chirp Doppler that appears between 30 and 120 s after the release may be resolved into two peaks with a separation of about 0.7 Hz. If the angle of arrival of the peak with the lower Doppler shift is calculated when it is relatively stable (60–100 s), the mode arrives from an azimuth of about 10° to the south of the nominal at an elevation angle of 28° . Although this is an extreme ray its azimuth does match the direction of the release, which makes the apparent direction physically plausible. The mode with a Doppler shift of -0.4 Hz (it reaches a juncture at about 2090 s) has an even greater azimuthal deflection than the high ray and would only reasonably be produced by a ray path deflected by the NICARE depletion.

The weak mode that moves off to increasingly negative Doppler shift after 900 s arrives from far to the south of the nominal path. This also is an extreme ray which must have been produced by the depletion. The relatively rapid change in direction of arrival is consistent with the high Doppler shift of the mode.

GROUP PATH CHANGES

If a single mode can be isolated then spaced-frequency CW transmissions can yield information

about group path changes. Essentially, one tries to synthesize a pulse transmission using a finite number of frequencies. Because the relationship of the relative phases of the transmissions is not known one cannot measure the absolute group path. That is, the waveform of a transmission consisting of two CW frequencies will consist of an amplitude modulation corresponding to the beat frequency. If the relative phase of the two frequencies is unknown then one cannot determine when the maximum amplitude occurred at the transmitter so that the timing is unknown to within a beat period. A pulse transmission overcomes this limitation by fixing the relative phases of the different frequencies. However, if the phases of the CW transmissions are fixed then the signal will consist of an amplitude modulation corresponding to a pseudopulse train and it thus should be possible to obtain the change in group path with time by observing the change in the modulation envelope. For two frequencies, one is limited in the total change in group path that is uniquely identifiable. With multiple frequencies, one has redundant information to set limits on the change.

Figure 7a shows the results for the group path change of the low ray on the Bermuda to Ohio path between 0 and 2700 s after launch; that is, we selected a frequency range of -0.3 to $+0.2$ Hz to isolate the major mode during that time interval. Group path information was obtained by the NRL channel probe during two segments of this interval: the first between 140 and 800 s and the second between 2360 and 3000 s. The group path change obtained by the CW technique is consistent with the Channel Probe results: a relatively stable delay during the first segment and a rapid increase in delay leading up to the junction with the high ray during the second. The group delay change between 2300 and 2700 s for the high ray derived using only the three lowest CW frequencies is shown in Figure 7b. The fourth frequency did not give consistent results. Figure 7b shows a decrease in group delay as the path approached the junction; this is consistent with the Channel Probe results. We were not able to estimate the difference in group delay between the low and high modes because the relative phase was not stable enough. Nor were we able to derive group delay change on the weaker modes in the spectra. The technique requires a high degree of coherence between the different frequencies in order to get a stable estimate of the phase.

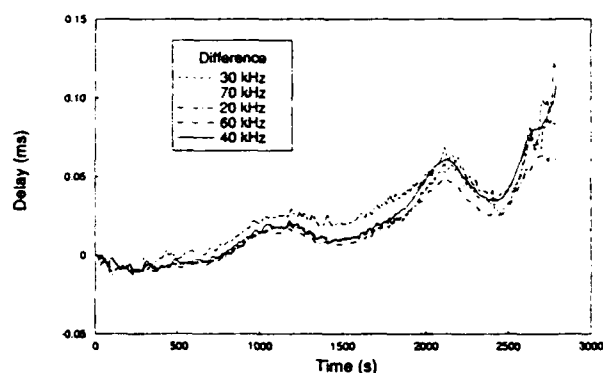


Fig. 7a

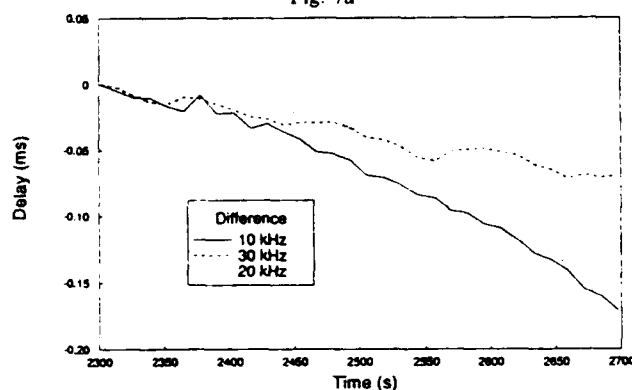


Fig. 7b

Fig. 7 Group path charges for the low ray on the Bermuda to Ohio path

FADING AND AMPLITUDE STATISTICS

Fading and amplitude effects on Florida to Nova Scotia path

The effect of the multipath produced by the depletion on the Florida to Nova Scotia path is graphically illustrated in Figure 8 where we display the total power received on one antenna during two time periods. The first, shown in Figure 8a, is between 500 and 100 s before the rocket launch while the second, shown in Figure 8b, is between 500 and 900 s after launch. The power was estimated in a 1 s time window with no overlap between successive windows. There is an obvious increase in fading rate in the data after the release. Of course, the increase in fading is another way of interpreting the broadening of the Doppler spectra after the release as shown in Figure 5 above. That is, the spectra initially had a width of about 0.1 Hz which increased to about 0.4 Hz after the release. The widths correspond to a decrease in the fading

period from about 10 to 2.5 s. The amplitude statistics before and after the release are similar and resemble Rayleigh fading.

Fading and amplitude statistics on Bermuda to Ohio path

Examination of the amplitude statistics on the Bermuda to Ohio path shows a distinctly different fading pattern than that described above. Figure 9 shows the received signal power for the time period between 900 and 1300 s after the rocket launch. This period is after the chemical releases but before the onset of strong multipath between the high and low rays. As compared to the fading observed on the Nova Scotia data (Figure 8) the received signal power is relatively steady and dominated by long periods. If one tries to compare the amplitude statistics with a Rayleigh distribution, one finds that the fit is very poor and is closer to a lognormal distribution in power.

FREQUENCY COHERENCE

We broadcasted four separate CW signals from two antennas. The frequencies of the transmissions were spaced at 0, 10, 30, and 70 kHz above the base frequency for both paths except that the fourth frequency was at 80 kHz for the Florida to Nova Scotia path. The use of these multiple CW transmissions enabled us to measure the coherence of the path, $\langle \psi^*(\Delta f) \psi(0) \rangle$, for a frequency separation Δf , where ψ is the complex amplitude of the received signal. We note that one effect of dispersion, the dependence of the index of refraction upon frequency, is that the phase and group paths differ, which from (4) indicates that the phase path and therefore the phase varies with frequency. This dependence may be considered as a linear and higher-order variation of the phase with frequency. The linear dependence gives the group path, while the higher-order dependence arises from the variation of the group path with frequency and may be observed as a pulse spreading. Since the initial phases of our transmissions were unknown we were not able to determine this dispersive phase shift, which would have allowed us to find the absolute group path. Nor are we sensitive to the nonlinear variation of the phase with frequency, so our results do not include the effect of pulse spreading due to

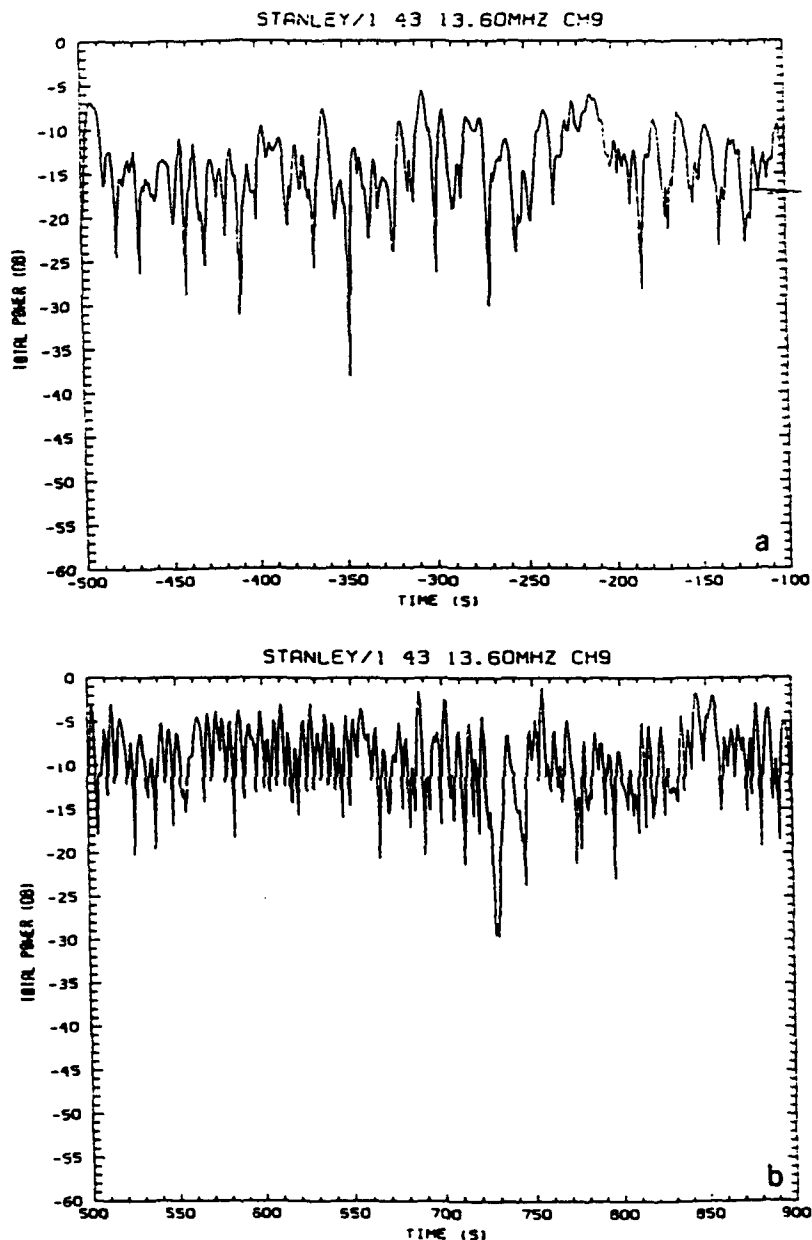


Fig. 8. Total power received on one antenna (Stanley Site) both (a) before and (b) after the chemical release. The increase in the fading rate after the release is due to the new propagation modes.

dispersion. Thus we are chiefly interested in the magnitude of the frequency coherence.

Figure 10 shows the frequency coherence of the Stanley data for separations of 10 and 50 kHz. The coherence was calculated using (12) with a time window of 100 s and averaged over the Doppler band from -0.5 to $+0.5$ Hz. The results indicate

that the initial coherence was degraded but that after the depletion was formed there was a marked decrease in coherence. The coherence on the Bermuda to Ohio path was high and showed slow changes arising from phase changes as the group delay of the one-hop mode changed. No loss in coherence analogous to that shown on the Nova

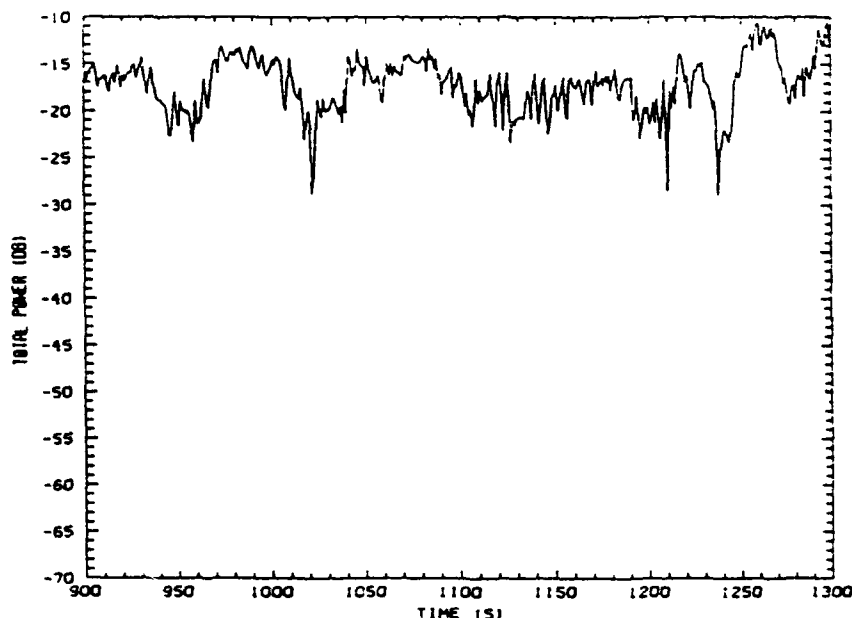


Fig. 9. Total power received for the Bermuda-Ohio path at 900–1300 s.

Scotia path was observed nor would be expected given the weakness of the depletion-induced modes.

MODELING THE NICARE EVENT AND BISTATIC PROPAGATION

Chirped signal

The bistatic Doppler data for both propagation paths showed highly chirped but weak negative

Doppler peaks associated with the second chemical release during the first 100 s. On the Bermuda to Warren path the signal disappeared while still showing a negative Doppler displacement from the main signal. On the Orlando to Stanley path the signal crossed through zero Doppler and became a “new” propagation mode that lasted for tens of minutes. The ray trace program TRACKER could not identify any “homed” modes in the first 100 s following the release, although after that it did find consistent modes for the Orlando to Stanley path. In fact, these chirped modes do come from the release, which at very early times acts like a very short focal length lens. During this period the depletion is strongly bending the rays that pass through it.

The depletion is 75 km to the north of the great circle path for the Bermuda to Ohio path and so might not be expected to normally affect propagation along this path. Figures 11a–11d show the modeled development of the depletion and its effect on propagation during the first 300 s. After the 100 s of growth the depletion can be seen to bend or focus the bundle of rays that pass through it. These rays are so strongly bent that some of them in fact do reach the receiver, far to the south of where they might nominally land. Because of the massive defocusing, the power in this new “mode” might be

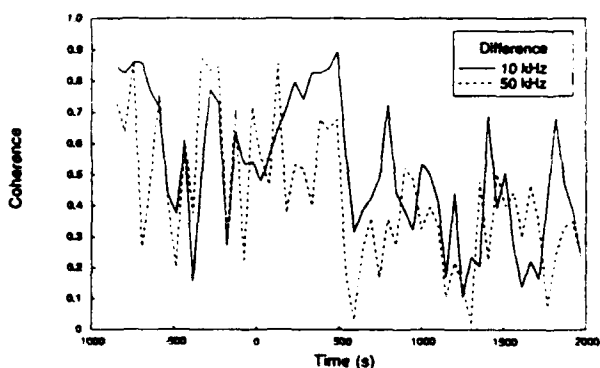


Fig. 10. The frequency coherence for the Nova Scotia data, for frequency separations of 10 and 50 kHz. Note that after the depletion there is a marked decrease in coherence.

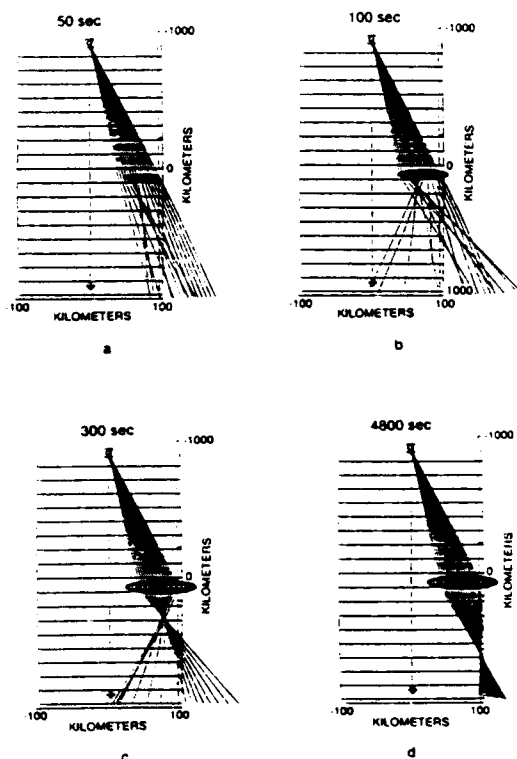


Fig. 11. The modeled effects of the depletion on a bundle of HF rays at 4 times during the depletion growth (50, 100, 300, 4800 s) for the Bermuda-Ohio path. Note the lenslike focusing effect of the depletion.

expected to be many tens of decibels down from the main mode. The rapidly changing focal length of the depletion formed "lens" is evidenced as rapidly changing phase and group paths, which accounts for the chirped Doppler seen in the bistatic data.

Figures 12a–12d show a similar development for the Orlando to Stanley path, except this time the depletion lies very nearly on the great circle path. Again, massive defocusing occurs initially, but after the "lens" has grown and weakened a identifiable mode is found to exist.

Long-lived modes

After the initial rapidly changing depletion chemistry the ray-tracing program (TRACKER) was unable to identify any new long-term propagation modes that would be generated on the Bermuda to Ohio path by the depletion, which was located nearly 75 km north of the great circle path. Figure 11 shows the geometry, indicating that the focal

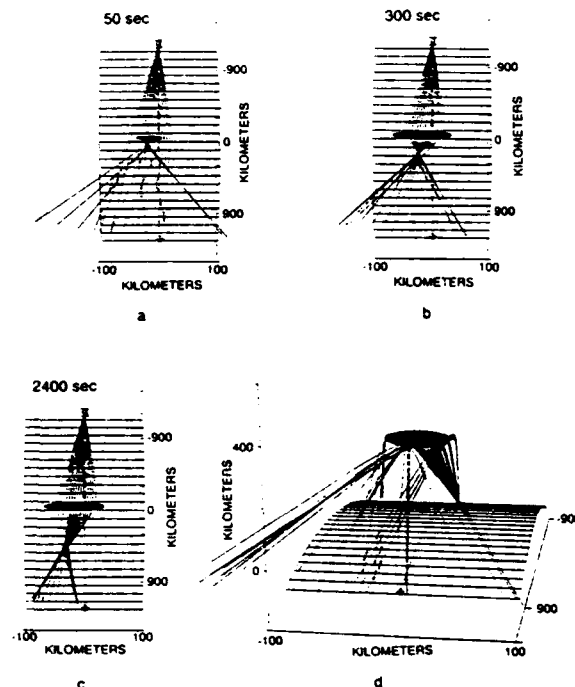


Fig. 12. The modeled effects of the depletion on the Florida to Nova Scotia path. Figure 30d shows a three-dimensional perspective.

length of the "lens" has grown long enough not to "scatter" rays near the receiver.

On the other hand, TRACKER was very successful identifying new paths generated on the Orlando to Stanley path, mainly because the depletion began almost directly on the midpoint of this path. Figure 13a shows the predicted mode for this path, just before the release. We could only find and identify the "low" ray. The bistatic data does not indicate that a strong second mode was present.

Between 50 and 100 s after release two new propagation modes had been formed. As Figure 13b shows, these new modes are directly related to the presence of the ionospheric depletion. At 600 s after release the model indicates the presence of two depletion related modes, see Figure 13c. Hence during these periods, homed rays were found that were refracted around the east and west sides of the depletion (Figures 13b and 13c), in agreement with the spaced receiver measurements.

As the depletion grows in size and fills in, the configuration of the original mode, and the "re-

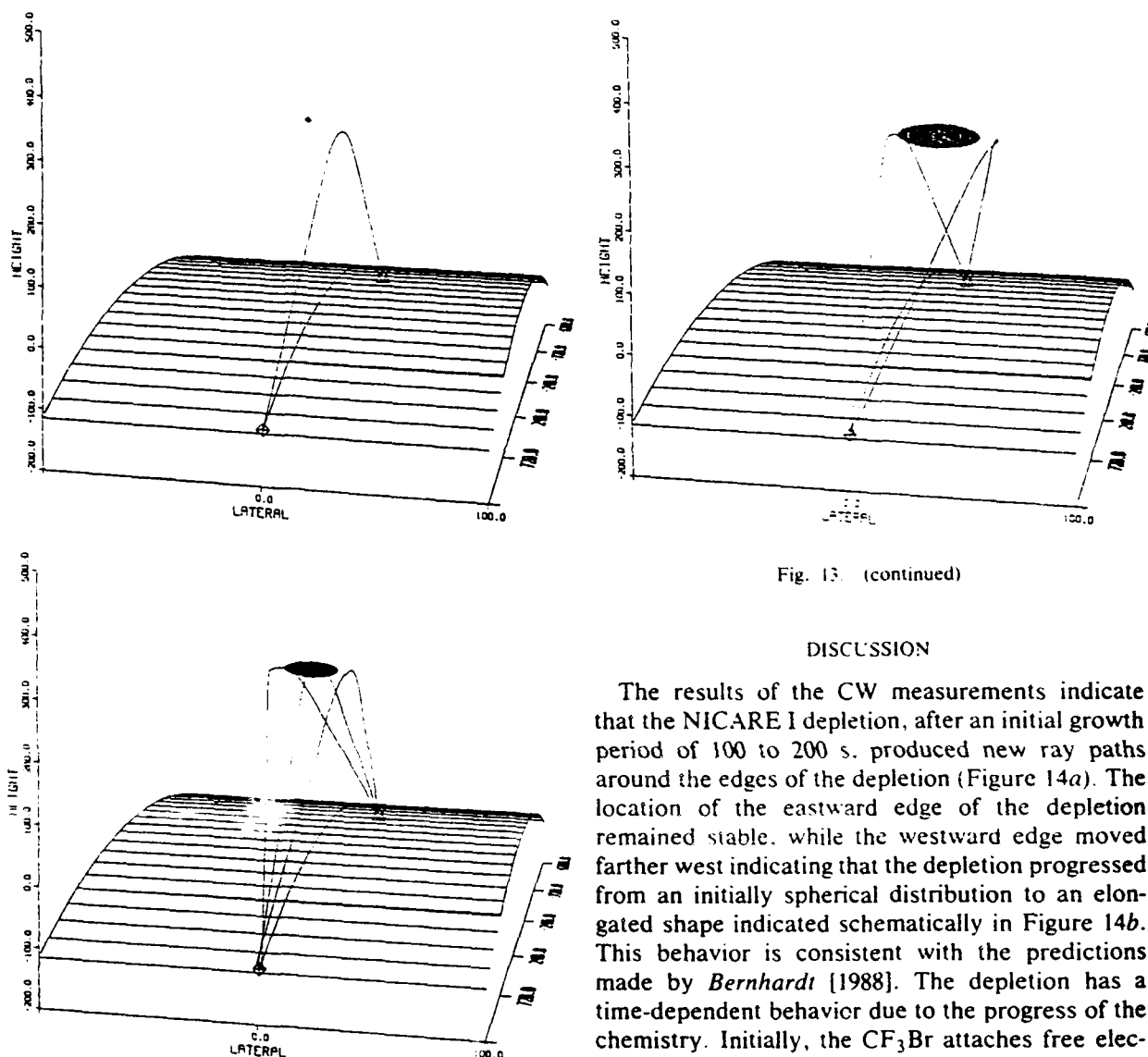


Fig. 13. (continued)

DISCUSSION

Fig. 13. "Homed" rays on the Florida-Nova Scotia path for prerelease and two different stages of depletion development. The prerelease model indicates only a low ray exists. Between 50 and 100 s after release two "depletion" modes (Figure 13b) showed up, by 600 s after release one mode disappeared (Figure 13c).

lease" modes changes. Figure 13c shows how the "release" modes have moved west, tracking the westward movement of the depletion. These model simulations have considered the depletion as drifting westward at 50 m/s.

The results of the CW measurements indicate that the NICARE I depletion, after an initial growth period of 100 to 200 s, produced new ray paths around the edges of the depletion (Figure 14a). The location of the eastward edge of the depletion remained stable, while the westward edge moved farther west indicating that the depletion progressed from an initially spherical distribution to an elongated shape indicated schematically in Figure 14b. This behavior is consistent with the predictions made by *Bernhardt* [1988]. The depletion has a time-dependent behavior due to the progress of the chemistry. Initially, the CF_3Br attaches free electrons to produce negative ions and an electron depletion. Assuming that the ambient electric field is zero and that the neutral wind is from the west, this ion cloud will polarize and produce an electric field which causes a plasma drift in the eastward direction. However, after about 100 s the CF_3Br^- will combine with the ambient O^+ to produce a neutralized volume. Polarization charges will gather on the edges of this volume but in the opposite sense as before; the plasma drift will be reversed, that is, antiparallel to the neutral wind. Moreover, the westward edge is unstable and may be expected to structure. Direction of arrival obtained from the signals received after penetration of the MUF indicate that they were scattered from the westward

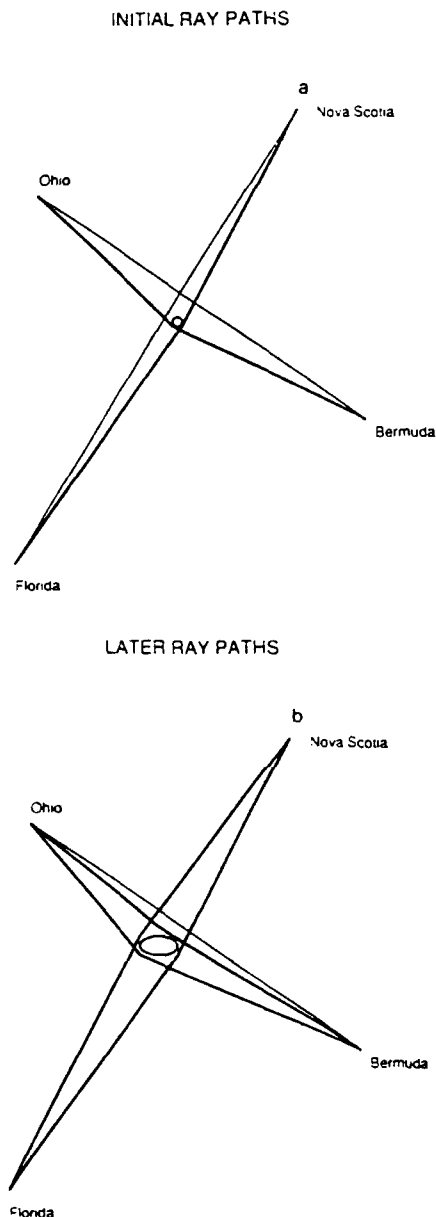


Fig. 14. Schematic view of the ray paths as measured (and modeled) for 2 times in the depletions growth.

edge of the depletion indicating that indeed there was structure on that edge.

APPENDIX A: A THREE-DIMENSIONAL HAMILTONIAN RAY-TRACING PROGRAM FOR IONOSPHERIC RADIO PROPAGATION

This computer program tracks (hence is named TRACKER) the three-dimensional paths of radio

waves through model ionospheres by numerically integrating Hamilton's equations, which are a differential expression of Fermat's principle. The Hamiltonian method, by using continuous models, avoids false caustics and discontinuous raypath properties often encountered in conventional ray-tracing methods. In addition to computing the ray path, TRACKER also calculates the group (or pulse travel time), the phase path, geometrical (or "real") path length, and Doppler shift (if the time variation of the ionosphere is explicitly included). Computational speed can be traded for accuracy by specifying the maximum allowable integration error per step in the integration.

This program is an extension of a three-dimensional Hamiltonian integration code developed in the late "1960s" by R. Michael Jones and J. M. Stephenson, commonly called the "Jones code" [Jones and Stephenson, 1975]. We have substituted a modern linear differential equation solving routine [Kahaner et al., 1989] for the Runge-Kutta solver in the Jones code. This not only allows the program to integrate accurately with larger step sizes but also enables the program to integrate through spatially steeper electron density gradients with smaller integration errors.

This program, TRACKER, has followed the "Jones' code" concept of modularity. The program can be extended to integrate any geophysical form of raytracing by simply replacing the subroutine defining the Hamiltonian. The models of the background index of refraction, and perturbations thereon, are also in a few replaceable subroutines.

TRACKER operates implicitly in several coordinate systems. The interface to the user is done in the usual geographic coordinates, but internally the actual ray path is calculated in the more physical geomagnetic coordinates (note here that we use a tilted dipole, rather than the more complex and accurate "corrected" geomagnetic coordinates). Because we are specifically interested in the effects of perturbations on the ray paths we also include coordinate systems relevant to each perturbation, making those calculations more tractable. TRACKER calculates gradients in the index of refraction explicitly, thus allowing nonanalytic forms of ionospheres and perturbations.

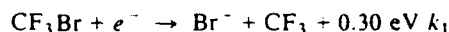
Ray-tracing codes cannot in general compute the ray path that connects a specified source and receiver. This "homing" capability is usually included by launching a fan of rays at small incre-

ments of azimuth and elevation, and linearly interpolating to find the ray that reaches the receiver location. In many cases, ray paths must be found that approach to within fractions of a wavelength of the receiver. TRACKER includes a formalism to treat this "homing" problem as a differential equation of its own. In this case the two-dimensional zero crossings of the rayfans are calculated (azimuth and elevation are treated jointly). These are solved sequentially and iteratively until a "homed" ray of sufficient accuracy is found.

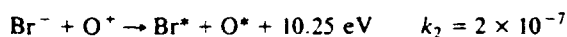
Unlike the original "Jones code," this program is not batch oriented but rather is interactive. When the program is started, the user then has approximately 75 commands with which to tailor his operating session. These commands cover an extensive range including controlling plot outputs (from none to directing it to special metafiles), modifying the physical ray trace parameters (such as frequency, receiver or transmitter locations, ray type, etc.), the numerical integration parameters (step size, error tolerances, etc.), and complex operations such as homing, finding multipath situations, etc.

APPENDIX B: IONOSPHERIC DEPLETION MODEL

The chemistry of neutral releases in the ionosphere is complex, with many reactions involving multiple ion species [Mendillo and Forbes, 1982; Bernhardt, 1987; Mendillo, 1988]. In the NICARE I experiment, 30 kg of trifluoromethyl bromide (CF_3Br) was released into the nighttime F layer (309 km). CF_3Br undergoes a relatively simple chemical reaction, first dissociatively attaching electrons to form Br^- , which then mutually neutralizes with O^+ .



$$= \frac{2.8 \times 10^{-7}}{1 + 0.6 \exp(998/T)}$$



See Bernhardt [1987] for a more detailed description of the chemistry. In particular, note the temperature dependence of the dissociative attachment rate, which varies from 6.7×10^{-8} at 600 K to 1.5×10^{-8} at 300 K.

We have included diffusion for the neutrals and ignored neutral winds and gravity. We have, however, included plasma transport via ambipolar diffusion.

In the initial model we expected the CF_3Br to be at roughly the background neutral temperature, say 600–800 K. The data from both the bistatic paths and the Wallops ionosonde indicate this would give too large a depletion. Our data is best fit by rate coefficients corresponding to a temperature nearer 300 K. Bernhardt (private communication, 1991) suggests that this is due to the adiabatic cooling of the CF_3Br as it expands in the near vacuum.

Acknowledgments. This work was supported by the U.S. Department of Energy. Cooperation of the Technical University of Nova Scotia is gratefully acknowledged. The Wallops Island ionograms were courtesy of the U. S. Air Force. The authors thank Dexter Sutherland and John Zinn for substantial help in developing the depletion model. The authors would also like to thank Abram Jacobson for many useful discussions.

REFERENCES

- Basler, R. P., G. H. Price, R. T. Tsunoda, and T. L. Wong, Ionospheric distortion of HF signals, *Radio Sci.*, 23, 569–579, 1988.
- Bernhardt, P. A., A critical comparison of ionospheric depletion chemicals, *J. Geophys. Res.*, 92, 4617–4628, 1987.
- Bernhardt, P. A., Cross-B convection of artificially-created negative-ion clouds and plasma depressions, *J. Geophys. Res.*, 93, 8696–8704, 1988.
- Booker, H. G., A local reduction of F -region ionization due to missile transit, *J. Geophys. Res.*, 66, 1073–1079, 1961.
- Budden, K. G., *The Propagation of Radio Waves*, Cambridge University Press, New York, 1985.
- David, F., A. G. Franco, H. Sherman, and L. B. Shucavage, Correlation measurements on an HF transmission link, *IEEE Trans. Commun.*, COM-17, 245–256, April 1969.
- Davies, K., *Ionospheric Radio Waves*, Blaisdell, Waltham, Mass., 1969.
- Helms, W., and A. Thompson, Ray-tracing simulation of ionization trough effects upon radio waves, *Radio Sci.*, 8, 1125–1132, 1973.
- Jones, R. M., and J. M. Stephenson, A versatile three-dimensional raytracing computer program for radio waves in the ionosphere, *OT Rep. 75-76*, U.S. Dep. of Comm., Off. of Telecommun., Boulder, Colo., 1975.
- Kahaner, D., C. Moler, and S. Nash, *Numerical Methods in Software*, chap. 8, Prentice-Hall, Englewood Cliffs, N. J., 1989.
- Mendillo, M., Ionospheric holes: A review of theory and recent experiments, *Adv. Space Res.*, 8, 51–62, 1988.
- Mendillo, M., and J. Forbes, Theory and observation of a

- dynamically evolving negative ion plasma, *J. Geophys. Res.*, 87, 8273-8285, 1982.
- Otnes, R. K., and L. Enochson, *Applied Time Series Analysis*, John Wiley, New York, 1978.
- Taheri, S. H., and B. D. Steinberg, A measurement system and analysis procedure for determining the spatial phase structure function of ionospherically reflected waves, *IEEE Trans. Antennas Propag.*, AP-27, 500-507, July 1979.
- Wagner, L. S., and J. A. Goldstein, Wideband HF channel prober: System description, *Rep. 8622*, Nav. Res. Lab., Washington, D. C., March 1983.
-
- P. Argo, R. Carlos, and T. Joseph, Los Alamos National Laboratory, Los Alamos, NM 87545.

Evolution of Gaussian beams in the nonuniform Earth-ionosphere waveguide

V. A. Baranov, A. L. Karpenko, and A. V. Popov

Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, Troitsk, Russia

(Received February 16, 1991; revised September 12, 1991; accepted September 17, 1991.)

In a uniform refraction waveguide the transversal width of a Gaussian beam oscillates with an approximately linearly growing amplitude. There exists a special set of initial data which eliminates its global diffraction divergency. We obtain an analogous result for a horizontally nonuniform Earth-ionosphere duct. By choosing initial elevation angle and operating frequency it is possible to concentrate a narrow radio wave beam at any given point of the Earth's surface. Our analysis is based on asymptotic integration of the ray equations which is outlined in appendix.

1. INTRODUCTION

The behavior of Gaussian beam propagating in a homogeneous medium or along the axis of a uniform waveguide is well known [Goncharenko, 1977]. In the former case, as a result of diffraction effects, a paraxial wave packet of initial cross-section d_0 diverges monotonously up to the angular width $\theta \sim \lambda/d_0$ in the far zone (here λ is the wavelength). In the latter case the beam has an oscillating width and a periodically reproducing amplitude. Since superpositions of Gaussian beams are convenient for representation of arbitrary wave fields [Popov, 1982], it is of interest to study their characteristics also in nonuniform media. Earlier approximate formulae for parameters of a Gaussian beam propagating in a slowly varying dielectric waveguide have been obtained [Karpenko and Popov, 1988]. It follows from Karpenko and Popov [1988] that generally the beam width in the waveguide is subject to quasi-periodic spatial oscillations with approximately linearly growing envelope. An exclusive example is the paraxial beam in a dielectric waveguide with quadratic profile of permittivity $\epsilon(Z) = a - bZ^2$. In this particular case there is no global divergency of the beam. It is of great practical interest to find nondivergent Gaussian beams in natural nonuniform waveguides.

In the paper to follow the results of [Karpenko and Popov, 1988] are used to study the long-range propagation of decameter radio waves. We consider

a Gaussian beam as a model of the wave field generated by a narrow-directional transmitting antenna. Using energy conservation law, one can visualize the amplitude distribution in the Earth-ionosphere waveguide caused by refraction and diffraction effects, just following the beam width evolution. It will be shown that in a model two-dimensional problem a Gaussian beam can be focused in an arbitrary point down to its primary width by trimming operating frequency and initial elevation angle. It means that the divergency caused by diffraction can be compensated, and, in practice, only absorption and wave scattering in ionosphere will contribute to the signal attenuation.

In our analysis, just for simplicity, we consider the two-dimensional scalar problem and neglect the anisotropy of ionospheric plasma. Because the refraction indices of the ordinary and extraordinary components in the lower part of the ionospheric layer are practically identical, Earth's magnetic field does not affect on trajectories of grazing incident waves. Thus, if one is not interested in polarization effects, the scalar approximation is adequate.

It is known that the parameters of Gaussian beams can be expressed in terms of geometric optics [Babič and Buldyrev, 1991; Popov, 1982]. In order to obtain analytic formulae we use the asymptotic solution of the ray equations for a waveguide with small longitudinal gradients of electron density [Baranov and Popov, 1980, 1981]. In the appendix based on the results of [Baranov and Popov [1980, 1981] we outline the process of their asymptotic integration.

Copyright 1992 by the American Geophysical Union.

Paper number 91RS02639.
0048-6604/92/91RS-02639\$08.00

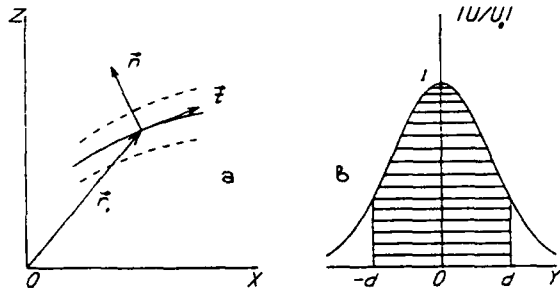


Fig. 1. Geometry of a Gaussian beam: (a) the central ray $r_0(p)$ with the tangent and normal vectors t and n ; (b) transversal field distribution $|U(Y, p)|/|U(0, p)|$.

2. ANALYSIS

Two-dimensional Gaussian beam in an isotropic nonuniform medium has the following field distribution [Popov, 1982]:

$$U = \frac{\text{const}}{\epsilon^{1/4}(p)[y_1(p) + \gamma y_2(p)]^{1/2}} \cdot \exp \left\{ ik \left[\Psi(p) + \frac{Y^2}{2} \Gamma(p) \right] \right\}. \quad (1)$$

Here $\epsilon(p)$ is dielectric permittivity $\epsilon(X, Z)$ taken on the central ray of the beam $r_0(p)$ —see Figure 1a; $p = \int \epsilon^{-1/2} ds$ and $\Psi = \int \epsilon^{1/2} ds$ are group path and eiconal along $r_0 = \{X(p), Z(p)\}$; $k = 2\pi/\lambda$ is wave number, and Y is the local coordinate transverse to the ray (along the normal vector n). Function $\Gamma(p) = d/dp \ln [y_1(p) + \gamma y_2(p)]$ with $\gamma = \mu + ix = \text{const}$ depends on some complete set $\{y_1(p), y_2(p)\}$ of transversal variations of the ray [Popov, 1982]. They can be found by differentiating the general solution $r_0(p, C_i)$ of the vector ray equation

$$\frac{d^2 \mathbf{r}}{dp^2} = \frac{1}{2} \nabla \epsilon(\mathbf{r}) \quad (2)$$

with respect to constants C_i and projecting onto the normal vector $n(p)$:

$$y_i(p) = \left(\mathbf{n}, \frac{d\mathbf{r}_0}{dC_i} \right). \quad (3)$$

It is obvious that for $\text{Im } \Gamma > 0$, (1) represents a beam with Gaussian amplitude distribution in its cross section (see Figure 1b). Its characteristic width $d(p)$ and local curvature radius of the wave front $\rho(p)$

can be expressed in terms of the imaginary and real parts of function $\Gamma(p)$:

$$d(p) = \sqrt{\frac{2}{k \text{Im } \Gamma}}, \quad \rho(p) = \frac{\sqrt{\epsilon(p)}}{\text{Re } \Gamma}. \quad (4a)$$

In order to get their explicit representation we introduce two real functions $\sigma(p) = (\text{Im } \Gamma)^{-1}$, $\chi(p) = (\text{Re } \Gamma)(\text{Im } \Gamma)^{-1}$ so that

$$d(p) = \sqrt{\frac{2}{k} \sigma(p)}, \quad \rho(p) = \frac{\sigma(p) \sqrt{\epsilon(p)}}{\chi(p)}. \quad (4b)$$

It can be easily shown that they obey a simple law of evolution along the axis ray $r_0(p)$:

$$\begin{aligned} \sigma(p) &= \sigma^{-1}(0)[1 + \chi^2(0)]A^2(p) + 2\chi(0)A(p)B(p) \\ &\quad + \sigma(0)B^2(p) \end{aligned} \quad (5)$$

$$\chi(p) = 0.5d\sigma/dp.$$

Here

$$A(p) = \frac{y_1(0)y_2(p) - y_2(0)y_1(p)}{y_1(0)y_2'(0) - y_1'(0)y_2(0)} \quad (6)$$

$$B(p) = \frac{y_2'(0)y_1(p) - y_1'(0)y_2(p)}{y_1(0)y_2'(0) - y_1'(0)y_2(0)}$$

form a new fundamental set of solutions of the ray variation equation normalized so that $A(0) = 0$, $B(0) = 1$, $A'(0) = 1$, $B'(0) = 0$; the symbol prime means d/dp . Provided that the function $r_0(p, C_i)$ and the initial values $d(0)$, $\rho(0)$ are given, formulae (4)–(6) enable one to track easily the Gaussian beam (1) evolution along its central ray.

Let us consider decimeter radio wave propagation between the Earth and a model ionospheric layer with small longitudinal gradients of electron density $N_e(X, Z) \text{ cm}^{-3}$ (see Figure 2a). Let $H \sim 300 \text{ km}$ be the characteristic height of the Earth-ionosphere waveguide and $L \sim 3000 \text{ km}$ be the longitudinal scale of the nonuniform dielectric permittivity distribution $\epsilon(X, Z) = 1 - 81 \cdot 10^{-6} N_e/f^2$ (here $f \text{ MHz}$ is the operating frequency). First of all, we shall find an approximate representation for the ray trajectories using asymptotic expansion of the solutions of the (2) in powers of the small parameter $\nu = H/L \ll 1$. For this aim we introduce dimensionless variables $\zeta = 1/\nu \ln(R + Z)/R$, $\varphi = X/R$, where

R is the Earth's radius. Then the ray equations (2) take the form

$$\nu^2 \frac{d^2 \zeta}{d\tau^2} = \frac{1}{2} \frac{\partial \bar{\epsilon}(\varphi, \zeta)}{\partial \zeta}, \quad \frac{d^2 \varphi}{d\tau^2} = \frac{1}{2} \frac{\partial \bar{\epsilon}(\varphi, \zeta)}{\partial \varphi} \quad (7)$$

containing the small parameter ν explicitly and equivalent to that considered in the appendix. Here $\bar{\epsilon}(\varphi, \zeta) = [1 + (Z/R)]^2 \epsilon(X, Z)$ is the modified dielectric permittivity with characteristic scales of order of unity with respect to variables (φ, ζ) , and new parameter along the ray $\tau = R \int (R + Z)^{-2} dp = R \int \epsilon^{-1/2} (R + Z)^{-2} ds$ has been introduced.

An asymptotic (with $\nu \rightarrow 0$) solution of (7) $r_0 = \{\zeta(\tau), \varphi(\tau)\}$ has been constructed by means of two-times method [Baranov and Popov, 1980]. It has the following form (see (A17)–(A18)):

$$\begin{aligned} & \pm \Omega \int_{\zeta^-}^{\zeta^+} \frac{d\zeta'}{(\bar{\epsilon} - Q)^{1/2}} + \nu \frac{\Omega}{Q^{1/2}} \int_{\zeta^-}^{\zeta^+} \frac{\bar{\epsilon} d\zeta'}{(\bar{\epsilon} - Q)^{3/2}} \frac{\partial}{\partial \bar{\epsilon}} \\ & \cdot \int_{\zeta^-}^{\zeta^+} (\bar{\epsilon} - Q)^{1/2} d\zeta'' - \nu Q^{1/2} \int_{\zeta^-}^{\zeta^+} \frac{d\zeta'}{(\bar{\epsilon} - Q)^{1/2}} \frac{\partial}{\partial \varphi} \\ & \cdot \left(\Omega \int_{\zeta^-}^{\zeta^+} \frac{d\zeta''}{(\bar{\epsilon} - Q)^{1/2}} \right) \frac{1}{\nu} = \int_0^\varphi \frac{\Omega}{Q^{1/2}} d\varphi' - 2\pi N - C, \end{aligned} \quad (8)$$

$$\tau = \int_0^\varphi \frac{d\varphi'}{Q^{1/2}} - \frac{\nu^2}{Q^{1/2}} \int_{\zeta^-}^{\zeta^+} \frac{d\zeta'}{(\bar{\epsilon} - Q)^{1/2}} \frac{\partial}{\partial \varphi} \int_{\zeta^-}^{\zeta^+} (\bar{\epsilon} - Q)^{1/2} d\zeta''.$$

Here $Q(\varphi, I)$ and $\Omega(\varphi, I) = -(\pi/2) \partial Q / \partial I$ are slowly varying functions determined by the adiabatic invariant conservation law (see appendix):

$$I = \int_{\zeta^-}^{\zeta^+} (\bar{\epsilon} - Q)^{1/2} d\zeta = \text{const.} \quad (9)$$

$\zeta^\pm = \zeta^\pm(\varphi, I)$ are turning points (solutions of equation $\bar{\epsilon}(\varphi, \zeta^\pm) = Q(\varphi, I)$), N is the number of hop or cycle of ray libration, $C = \text{const}$ —its initial phase. Using an appropriate model profile of $\bar{\epsilon}$, one may extract from (8) a more explicit approximate solution for the ray trajectory $\varphi(\tau), \zeta(\tau)$ containing two arbitrary constant I and C . Making use of (3) and substituting $r_0 = \{\varphi(\tau), \zeta(\tau)\}$, $C_1 = C$, $C_2 = I$, we get a fundamental set of solutions of the equation of transverse ray variation:

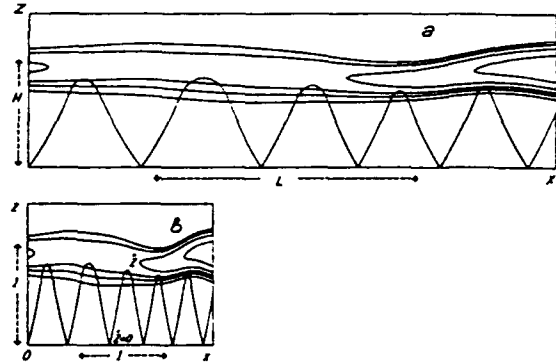


Fig. 2. Example of ray trajectory in a nonuniform waveguide. (a) the ray and lines of equal electron density N_e with the characteristic scales L and H ; (b) the same in normalized variables $x = X/L$, $z = Z/H$.

$$y_1(\tau) = \frac{\frac{\partial \varphi}{\partial \zeta} \frac{\partial \zeta}{\partial \tau} - \frac{\partial \zeta}{\partial \varphi} \frac{\partial \varphi}{\partial \tau}}{[\bar{\epsilon}(\varphi, \zeta)]^{1/2}}, \quad y_2(\tau) = \frac{\frac{\partial \varphi}{\partial I} \frac{\partial \zeta}{\partial \tau} - \frac{\partial \zeta}{\partial I} \frac{\partial \varphi}{\partial \tau}}{[\bar{\epsilon}(\varphi, \zeta)]^{1/2}} \quad (10)$$

Now we employ the approximate solution (8) for calculation of the partial derivatives in (10) and eliminate parameter τ . Then we obtain an explicit representation of transverse variations (10):

$$\begin{aligned} y_1 & \approx \frac{1}{\Omega} \left(\frac{Q}{\bar{\epsilon}} \right)^{1/2} \left[\pm (\bar{\epsilon} - Q)^{1/2} - \nu \frac{Q^{1/2} (\bar{\epsilon} - Q)^{1/2}}{\Omega} \frac{\partial}{\partial \varphi} \right. \\ & \cdot \left(\Omega \int_{\zeta^-}^{\zeta^+} \frac{d\zeta'}{(\bar{\epsilon} - Q)^{1/2}} \right) - \nu \left(\frac{Q}{\bar{\epsilon} - Q} \right)^{1/2} \frac{\partial}{\partial \varphi} \\ & \cdot \left. \int_{\zeta^-}^{\zeta^+} (\bar{\epsilon} - Q)^{1/2} d\zeta' \right] \end{aligned} \quad (11)$$

$$y_2 \approx \left[\frac{y_1}{\nu} \frac{\partial}{\partial I} \int_0^\varphi \frac{\Omega}{Q^{1/2}} d\varphi' + \frac{(\bar{\epsilon} - Q)^{1/2}}{\Omega} \frac{\partial}{\partial I} \cdot \left(\Omega \int_{\zeta^-}^{\zeta^+} \frac{d\zeta'}{(\bar{\epsilon} - Q)^{1/2}} \right) \right]$$

We used approximation (11) for calculating the transversal width $d(\tau)$ of a Gaussian beam of radio waves in the Earth-ionosphere duct via (4)–(6). Upon neglecting quantities of order $O(\nu)$ in (11) and returning to the original variables (X, Z) , one gets

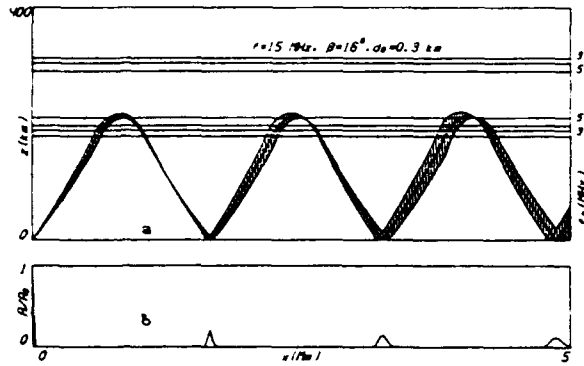


Fig. 3. Gaussian beam in a model uniform waveguide (general case): (a) the beam width and lines of equal plasma frequency $f_0 = 9 \times 10^{-3} N_e^{1/2}$ MHz; (b) relative amplitude distribution along Earth's surface.

$$A \approx \left(\frac{Q Q_0}{\tilde{\epsilon} \tilde{\epsilon}_0} \right)^{1/2} \left(\frac{v_0}{\Omega_0} u - \frac{u_0}{\Omega_0} v - u_0 u M \right) \quad (12)$$

$$B \approx \frac{u}{u_0} \left(\frac{Q \tilde{\epsilon}_0}{Q_0 \tilde{\epsilon}} \right)^{1/2} + \frac{Q_0}{\Omega_0 u_0 \tilde{\epsilon}_0} A$$

where

$$u(X, Z) = \pm \frac{(\tilde{\epsilon} - Q)^{1/2}}{\Omega}$$

$$v(X, Z) = \pi(\tilde{\epsilon} - Q)^{1/2} \frac{\partial}{\partial l} \left(\Omega \int_Z^{Z'} \frac{dZ'}{(R + Z')(\tilde{\epsilon} - Q)^{1/2}} \right)$$

$$M(X, l) = \frac{\pi}{R} \frac{\partial}{\partial l} \int_0^X \frac{\Omega}{Q^{1/2}} dX' \quad (13)$$

$$l = \int_Z^{Z'} (\tilde{\epsilon} - Q)^{1/2} \frac{dZ}{R + Z}$$

$$\frac{\pi}{\Omega} = \int_Z^{Z'} \frac{dZ}{(R + Z)(\tilde{\epsilon} - Q)^{1/2}}$$

In (12) we imply that functions $u(X, Z)$, $v(X, Z)$, $\tilde{\epsilon}(X, Z)$ are taken on the central ray $Z = Z(X, l, C)$ determined via (8) and (9), and those with subscript zero are calculated at the initial point $X = 0$, $Z = Z(0, l, C)$.

3 NUMERICAL RESULTS

Some typical examples of Gaussian beams of short radio waves calculated via (4)–(6) and (12) and (13) are presented in the Figures 3a–6a. The operating fre-

quency is chosen in the middle of the decimeter range: $f = 15$ MHz. The environment is shown with the lines of equal plasma frequency $f_0 = 9 \times 10^{-3} N_e^{1/2}(X, Z)$ MHz. The shaded zones depict the beam truncated at the lateral width $\pm d(\tau)$ (see Figure 1b). Because of contracting the picture along the X axes, the normal vector $\mathbf{n}(\tau)$ coinciding with the direction of hatching does not look to be orthogonal to the central ray. Nevertheless, the beam is depicted in the true scale, and its width can be measured directly along both X and Z axes. The Figures 3b–6b illustrate the normalized amplitude distribution of the Gaussian beam (1) on the Earth's surface.

The Figures 3 and 4 correspond to the model case of a spherically layered ionosphere with quasiparabolic profile of electron density $N_e(Z)$. Figure 3a illustrates a typical situation, the beam width oscillates with increasing swing. Note that for the initial value $d_0 = 0.3$ km the maximum width over the three-hop path ($L \sim 5$ Mm) is about several tens of kilometers. The behavior of the wave packet can be understood easily by using formulae (12) and (13). In the case of a uniform waveguide at hand, functions u and v are periodic in X and integral $M(X, l)$ is growing linearly. At great distances from the source the divergency rate of the beam, determined by the factor $M(X, l)$, is proportional to the derivative $\partial/\partial l(\Omega Q^{-1/2})$.

Formulae (12) and (13) allow to find a purely periodic regime of the Gaussian beam evolution. It can be shown that the product $\Omega Q^{-1/2}$ is reciprocally proportional to the hop length $D(\beta)$ which depends on the initial elevation angle $\beta = \cos^{-1}(Q^{1/2})$ of the center ray. Recalling that the function $D(\beta)$ has a minimum [Davies, 1966, p. 226], we see

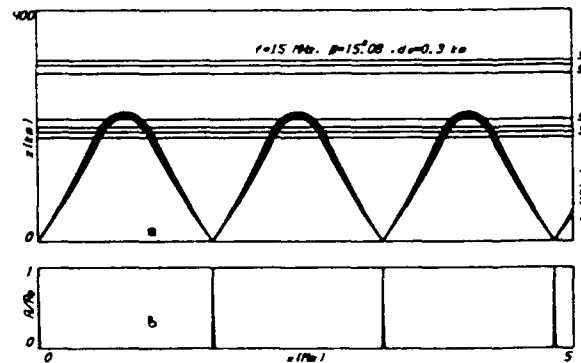


Fig. 4. Gaussian beam in a model uniform waveguide (periodical regime).

that, for the corresponding value of β , the factor $M(X, I) = (\pi X/R)(\partial/\partial I)(\Omega Q^{-1/2})$ equals to zero identically which eliminates the global divergency of the beam. This very case is depicted in Figure 4a. The periodic regime of the beam width evolution leads to quite a different pattern of field distribution along Earth's surface. In the general case (Figure 3b) the illuminated patches become wider from hop to hop which leads to gradual fading of the maximum signal amplitude. In the periodic case, only narrow regions of the order of $d/\sin \beta$ are illuminated, and successive amplitude maxima (Figure 4b) do not diminish with distance. (in practice, of course, one has to take into account absorption and scattering in the ionosphere and also the horizontal divergency of the beam leading to additional fading proportional to $[\sin X/R]^{-1/2}$.)

Another effect to be taken into account in practical calculations is regular variations of the ionospheric duct parameters along the long-range propagation path. Its influence may be calculated easily within the framework of the approximation discussed above. For the horizontally nonuniform Earth-ionosphere waveguide, function $M(X, I)$ cannot vanish identically. None the less, for a fixed path length $X = L$, the initial elevation angle β of a Gaussian beam can be chosen so that $M(L, I) = 0$ which provides maximum focusing in the given cross section of the duct.

Examples of evolution of Gaussian beams in the ionosphere are depicted in Figures 5 and 6. A quasiparabolic profile of F_2 layer has been utilized, with parameters taken from the long-term ionospheric forecast [Nosova and Chernyshev, 1981]. Figure 5a illustrates the common situation, and Figure 6a corresponds to such a choice of initial data of the beam

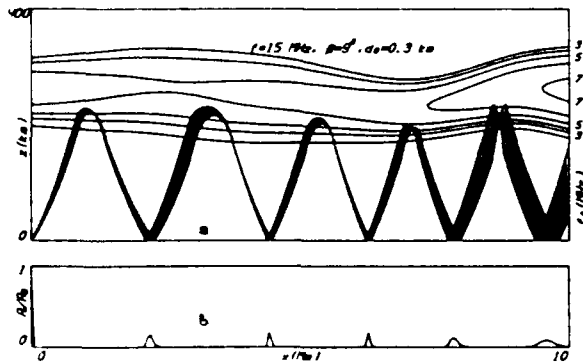


Fig. 5. Gaussian beam in an Earth-ionosphere duct (general case).

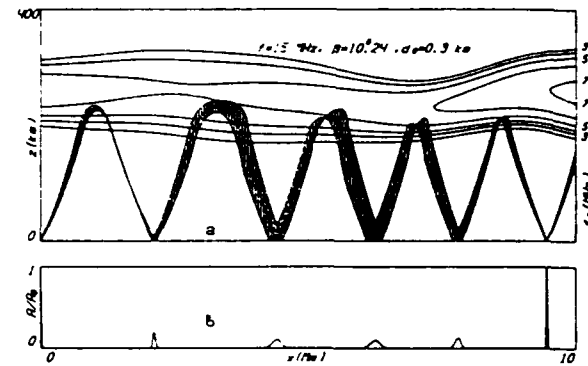


Fig. 6. Gaussian beam in an Earth-ionosphere duct (focusing at the given point).

that provides maximum focusing at the given point of Earth's surface. The field amplitude distribution along the propagation path is depicted in Figures 5b and 6b. It is seen that regular horizontal gradients in ionosphere do not destroy the beam. That justifies the use of the proposed approach for estimating the wave field generated by a narrow-directional transmitter over a long path $L \sim 10$ Mm.

4. CONCLUSION

To conclude, we shall show how the former criterion of focusing is connected with the pattern of caustics caused by a point source within the Earth-ionosphere wave duct. According to [Karpenko and Popov [1988], the caustic (the line where the divergency of a fan of ray trajectories $\zeta = \zeta[\varphi, I, C(I)]$ bound by the initial condition $\zeta[0, I, C(I)] = 0$ vanishes) is determined by the following equation:

$$\begin{aligned} \Re(\varphi, \zeta; I) &= \left| \frac{\partial \varphi}{\partial I} \frac{\partial \zeta}{\partial \tau} - \frac{\partial \zeta}{\partial I} \frac{\partial \varphi}{\partial \tau} \right| \approx \frac{Q^{1/2}}{\Omega} \\ &\cdot \left[\pm \frac{(\bar{\epsilon} - Q)^{1/2}}{v} + \left(\frac{Q}{\bar{\epsilon} - Q} \right)^{1/2} \frac{\partial}{\partial \varphi} \int_{\zeta}^{\zeta'} (\bar{\epsilon} - Q)^{1/2} d\zeta' \right. \\ &\quad \left. - \frac{Q^{1/2}(\bar{\epsilon} - Q)^{1/2}}{\Omega} \frac{\partial}{\partial \varphi} \left(\Omega \int_{\zeta}^{\zeta'} \frac{d\zeta'}{(\bar{\epsilon} - Q)^{1/2}} \right) \right] \\ &\cdot \frac{\partial}{\partial I} \int_0^{\varphi} \frac{\Omega}{Q^{1/2}} d\varphi' + (\bar{\epsilon} - Q)^{1/2} \\ &\cdot \left[\frac{\partial}{\partial I} \left(\Omega \int_{\zeta}^{\zeta'} \frac{d\zeta'}{(\bar{\epsilon} - Q)^{1/2}} \right) \pm \frac{\partial C}{\partial I} \right] = 0 \end{aligned} \quad (14)$$

together with the ray formula (8). For a surface-located source $X = 0$, $Z = 0$, we obtain $C = -\pi$, $\partial C/\partial l = 0$, and the condition that the ray is tangent to a caustic at the receiving point $X = L$, $Z = 0$, after returning from (φ, ζ) to (X, Z) takes the following form:

$$\mathfrak{B}(L, 0; l) = \frac{Q^{1/2}(\bar{\epsilon} - Q)^{1/2}}{H\Omega} \frac{\partial}{\partial l} \int_0^L \frac{\Omega}{Q^{1/2}} dX = 0$$

$$\frac{1}{H} \int_0^L \frac{\Omega}{Q^{1/2}} dX = 2\pi N, \quad (15)$$

(Here the adiabatic invariant conservation law (9) and resulting formula for $\Omega(X; l)$ have been used.) The first of these two conditions, evidently, coincides with the criterion of the focusing the Gaussian beam $M(L; l) = 0$, and the second one means that its center ray $X(\tau)$, $Z(\tau)$ comes to the observation point $X = L$, $Z = 0$. Together the two conditions (15) determine such a combination of the operating frequency f (upon which the dielectric permittivity of the ionosphere depends) and the initial elevation angle β of the Gaussian beam that provides maximum energy concentration at the given point. They may be useful for optimizing the transmitting regime of radio communication circuits.

APPENDIX

Here we give a brief outline of asymptotic ray tracing in a waveguide with small longitudinal gradients of dielectric permittivity. Let the vertical profile $\epsilon(Z/H)$ form a wave duct of characteristic width H and changes gradually along the horizontal axis X with characteristic scale L : $\epsilon = \epsilon(X/L, Z/H)$. In order to construct an approximate solution of the ray equations valid at long distances $X \sim L$ one has to separate fast oscillations of the ray due to the strong vertical gradients of ϵ and its slow evolution caused by the longitudinal nonuniformity. For this purpose, we introduce normalized components of the vector $\mathbf{r}_0(\tau)$: $x = X/L$, $z = Z/H$ (see Figure 2). As a result, (2) takes the form

$$\nu^2 \frac{d^2 z}{d\tau^2} = \frac{1}{2} \frac{\partial \epsilon(x, z)}{\partial z}, \quad \frac{d^2 x}{d\tau^2} = \frac{1}{2} \frac{\partial \epsilon(x, z)}{\partial x} \quad (A1)$$

containing small parameter $\nu = H/L \ll 1$. Then in accordance with the two-times method [Cole, 1968]

we seek for its solution as function of parameter τ and a new fast variable $t = 1/\nu \int \omega(\tau) d\tau$:

$$x(\tau) = \mathfrak{X}(t, \tau), \quad z(\tau) = \mathcal{Z}(t, \tau) \quad (A2)$$

where function $\omega(\tau)$ is to be found afterward. Introduction of the second variable transforms the singularly perturbed system (A1) to the form that allows constructing the solution as an asymptotic series in powers of ν :

$$\mathfrak{X}(t, \tau) = \mathfrak{X}_0 + \nu^2 \mathfrak{X}_2 + \nu^3 \mathfrak{X}_3 + \dots, \quad (A3)$$

$$\mathcal{Z}(t, \tau) = \mathcal{Z}_0 + \nu \mathcal{Z}_1 + \nu^2 \mathcal{Z}_2 + \dots,$$

(it can be shown that, without loss of generality, the term $\nu \mathfrak{X}_1$ can be omitted [Baranov and Popov, 1981]). Substituting (A3) into (A1) yields a recurrent set of differential equations for $\mathfrak{X}_i(t, \tau)$, $\mathcal{Z}_i(t, \tau)$:

$$\omega^2 \frac{\partial^2 \mathfrak{X}_0}{\partial t^2} = 0, \quad \omega^2 \frac{\partial^2 \mathfrak{X}_2}{\partial t^2} + \frac{\partial^2 \mathfrak{X}_0}{\partial \tau^2} = \frac{1}{2} \frac{\partial}{\partial x} \epsilon(\mathfrak{X}_0, \mathcal{Z}_0), \dots, \quad (A4)$$

$$\omega^2 \frac{\partial^2 \mathcal{Z}_0}{\partial t^2} = \frac{1}{2} \frac{\partial}{\partial z} \epsilon(\mathfrak{X}_0, \mathcal{Z}_0) \quad (A5)$$

$$\omega^2 \frac{\partial^2 \mathcal{Z}_1}{\partial t^2} + 2\omega \frac{\partial^2 \mathcal{Z}_0}{\partial t \partial \tau} + \omega' \frac{\partial \mathcal{Z}_0}{\partial t} = \frac{1}{2} \mathcal{Z}_1 \frac{\partial^2 \epsilon}{\partial z^2}, \dots,$$

here $\omega' = d\omega/d\tau$.

It can be seen easily that (A4)–(A5) can be solved subsequently as ordinary differential equations with respect to the fast variable t . The integration constants arising in the process of constructing \mathfrak{X}_i , \mathcal{Z}_i are actually functions of "slow time" τ and can be chosen in such a way that eliminates the secular (growing with the "fast time" t) terms in asymptotic series (A3).

First of all, we see that bounded with $t \rightarrow \infty$ solutions of the first equation of (A4) are functions of the sole slow variable τ : $\mathfrak{X}_0 = \mathfrak{X}_0(\tau)$. It enables us to integrate the first equation (A5),

$$t = \pm \omega(\tau) \int \frac{d\mathcal{Z}_0}{\sqrt{\epsilon[\mathfrak{X}_0(\tau), \mathcal{Z}_0] - q(\tau)}} + C(\tau) \quad (A6)$$

Inversion of the integral and matching the two branches corresponding to \pm signs in (A6) determine the zero-order approximation for the vertical component of the ray as periodic (and therefore bounded) function of t : $\mathcal{Z}_0(t, \tau) = F[t - C(\tau), \tau]$. Slowly varying functions $q(\tau)$ and $C(\tau)$ ("energy"

and shift constant) are to be found, like $\omega(\tau)$ and $\mathcal{X}_0(\tau)$, from the condition of boundedness of the corrections $\mathcal{X}_2, \mathcal{X}_1$. If we make the period of F constant: $T(\tau) = 2\pi$ by choosing

$$\omega(\tau) = \pi \left(\int_{z^-}^{z^+} \frac{dz}{\sqrt{\varepsilon[\mathcal{X}_0(\tau), z] - q(\tau)}} \right)^{-1} \quad (\text{A7})$$

where $z^+(\tau)$ and $z^-(\tau)$ are the turning points (roots of the equation $\varepsilon[\mathcal{X}_0(\tau), z] = q(\tau)$) (see Figure 2), all the right-hand sides in (A4) and (A5) will become periodical in t , which makes possible to formulate simple conditions of getting rid of secular terms in solutions $\mathcal{X}_2(t, \tau), \mathcal{L}_1(t, \tau)$.

These functions can be found from (A4) and (A5) as quadratures

$$\begin{aligned} \mathcal{X}_2(t, \tau) = & \frac{1}{\omega^2} \int_C^t dt' \int_C^{t'} \left(\frac{1}{2} \frac{\partial \varepsilon}{\partial x} (\mathcal{X}_0, \mathcal{L}_0) \right. \\ & \left. - \mathcal{X}_0'(\tau) \right) dt'' + \mathcal{X}_2(\tau) \end{aligned} \quad (\text{A8})$$

$$\begin{aligned} \mathcal{L}_1(t, \tau) = & -\frac{1}{\omega^2} \frac{\partial \mathcal{L}_0}{\partial t} \\ & \cdot \int_C^t \left(\frac{\partial \mathcal{L}_0}{\partial t} \right)^{-2} dt' \int_C^{t'} \frac{\partial}{\partial \tau} \omega \left(\frac{\partial \mathcal{L}_0}{\partial t} \right)^2 dt'' \\ & + \frac{\partial \mathcal{L}_0}{\partial t} \left[a_1(\tau) + b_1(\tau) \int_C^t \left(\frac{\partial \mathcal{L}_0}{\partial t} \right)^{-2} dt' \right] \end{aligned} \quad (\text{A9})$$

Because the integrand in (A8) contains the periodic function $\mathcal{L}_0(t, \tau)$, the inner integral is, in general, a sum of periodic and linear functions of the fast variable t . In order to eliminate the linear term we have to demand

$$\begin{aligned} \mathcal{X}_0'(\tau) = & \frac{1}{2} \frac{\partial \varepsilon}{\partial x} (\mathcal{X}_0, \mathcal{L}_0) \\ = & \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \varepsilon}{\partial x} [\mathcal{X}_0(\tau), \mathcal{L}_0(t, \tau)] dt \end{aligned} \quad (\text{A10})$$

which gives an equation for the slowly varying function $\mathcal{X}_0(\tau)$.

To accomplish constructing the zero-order approximation $\mathcal{X}_0, \mathcal{L}_0$, we have but to find the functions $q(\tau)$ and $C(\tau)$. For this purpose, let us see how

the secular terms in (A9) arise. The integrand of the first integral can be rewritten in the following form:

$$\omega C'(\tau) + \left(\frac{\partial \mathcal{L}_0}{\partial \tau} \right)^{-2} \frac{\partial}{\partial t} \int_C^t \omega \left(\frac{\partial \mathcal{L}_0}{\partial t} \right)^2 dt'' \quad (\text{A11})$$

Here in general, the integral may contain a linear in t term. In order to eliminate it we have to put the condition

$$\frac{\partial}{\partial \tau} \int_{-\pi}^{\pi} \omega \left(\frac{\partial \mathcal{L}_0}{\partial t} \right)^2 dt = 0 \quad (\text{A12})$$

Making use of (A6), one can transform it into the equivalent form

$$\frac{\partial I}{\partial \tau} = \frac{\partial}{\partial \tau} \int_{z^-}^{z^+} \sqrt{\varepsilon[\mathcal{X}_0(\tau), z] - q(\tau)} dz = 0 \quad (\text{A13})$$

which means that the integral I is constant along the ray (adiabatic invariant). Equation (A13) enables us to determine the slowly varying "energy" $q(\tau)$ as follows. Let us define two functions $Q(x) = Q(x, I)$, $\Omega(x) = \Omega(x, I)$ depending on the variable x and the invariant I via

$$I = \int_{z^-}^{z^+} \sqrt{\varepsilon(x, z) - Q(x)} dz = \text{const} \quad (\text{A14})$$

$$\Omega(x) = \pi \left(\int_{z^-}^{z^+} \frac{dz}{\sqrt{\varepsilon(x, z) - Q(x)}} \right)^{-1}$$

Then it proves that $q(\tau) = Q[\mathcal{X}_0(\tau), I]$, $\omega(\tau) = \Omega[\mathcal{X}_0(\tau), I] = -(\pi/2) \partial Q / \partial I$, and (A10) takes the equivalent form,

$$\mathcal{X}_0'(\tau) = \frac{1}{2} \frac{\partial Q}{\partial x} (\mathcal{X}_0, I), \quad \mathcal{X}_0'(\tau) = Q^{1/2}(\mathcal{X}_0, I) \quad (\text{A15})$$

separated from the vertical component $\mathcal{L}_0(t, \tau)$. Integrating (A15) with the initial condition $\mathcal{X}_0(0) = 0$, one obtain

$$\tau = \int_0^{\mathcal{X}_0} \frac{dx}{Q^{1/2}(x, I)} \quad (\text{A16})$$

The second integration in (A9) gives rise to another linear term unless we put $C'(\tau) \equiv 0$. So we have determined all the slowly varying functions in the zero-order approximation $\mathcal{X}_0(\tau), \mathcal{L}_0(t, \tau)$ and have got an approximate solution containing two

arbitrary constants: I and C . In similar way the functions $a_1(\tau)$, $b_1(\tau)$, and $\xi_2(\tau)$ in (A8) and (A9) can be found from the condition of boundedness of the following terms of asymptotic series (A3).

Now, when the zero-order approximation $\mathcal{L}_0(\tau)$, $\mathcal{L}_0(t, \tau)$ and the corrections $\mathcal{L}_2(t, \tau)$, $\mathcal{L}_1(t, \tau)$ are found, we can substitute them into (A3) and exclude the parameter τ by means of (A8) and (A16):

$$\tau \approx \int_0^x \frac{dx'}{Q^{1/2}} - \frac{\nu^2}{Q^{1/2}} \int_z^{z'} \frac{dz'}{(\varepsilon - Q)^{1/2}} \frac{\partial}{\partial x} \int_{z'}^{z''} (\varepsilon - Q)^{1/2} dz'' \quad (A17)$$

In this way we obtain an explicit approximate representation of the ray trajectory $z = z(x, I, C)$

$$\begin{aligned} & \pm \Omega \int_z^{z'} \frac{dz'}{(\varepsilon - Q)^{1/2}} \\ & + \nu \frac{\Omega}{Q^{1/2}} \int_z^{z'} \frac{\varepsilon dz'}{(\varepsilon - Q)^{3/2}} \frac{\partial}{\partial x} \int_{z'}^{z''} (\varepsilon - Q)^{1/2} dz'' \\ & - \nu Q^{1/2} \int_z^{z'} \frac{dz'}{(\varepsilon - Q)^{1/2}} \frac{\partial}{\partial x} \left(\Omega \int_{z'}^{z''} \frac{dz''}{(\varepsilon - Q)^{1/2}} \right) \\ & = \frac{1}{\nu} \int_0^x \frac{\Omega}{Q^{1/2}} dx' - 2\pi N - C. \end{aligned} \quad (A18)$$

which has been used above for calculation of the parameters of Gaussian beams.

REFERENCES

- Babič, V. M., and V. S. Buldyrev, *Short-Wavelength Diffraction Theory (Asymptotic Methods)*, Springer-Verlag, New York, 1991.
- Baranov, V. A., and A. V. Popov, Asymptotic integration of the ray equations in a three-dimensional almost stratified medium, Decimeter radio wave propagation (in Russian), *IZMIRAN Acad. Sci. USSR*, 167-171, 1980.
- Baranov, V. A., and A. V. Popov, Corrections to the adiabatic approximation for rays in a refraction waveguide, *Waves and diffraction (in Russian)*, *Acad. Sci. USSR*, 3, 199-202, 1981.
- Cole, J. D., *Perturbation Methods in Applied Mathematics*, Blaisdell, Waltham, Mass., 1968.
- Davies, K., *Ionospheric Radio Propagation*, Dover, New York, 1966.
- Goncharenko, A. M., *Gaussian Beams of Light (in Russian)*, *Nauka Tekh. Minsk, Russia*, 1977.
- Karpenko, A. L., and A. V. Popov, Caustics and Gaussian beams in a nonuniform refraction waveguide, *Radiotekh. Elektron.*, Moscow, 23(6), 1126-1131, 1988.
- Nosova, G. N., and O. V. Chernyshev, Algorithm and program for calculation of some characteristics of short radio wave propagation (in Russian), *IZMIRAN Acad. Sci. USSR*, 1981.
- Popov, M. M., A new method of computation of wave fields using gaussian beams, *Wave Motion*, 4(1), 85-97, 1982.
- V. A. Baranov, A. L. Karpenko, and A. V. Popov, Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, 142092, Troitsk, Russia.

Radiation of whistler waves in magnetoactive plasma

I. G. Kondrat'ev and A. V. Kudrin

Department of Radiophysics, University of Nizhny Novgorod, Nizhny Novgorod, Russia

T. M. Zaboronkova

Radiophysical Research Institute, Nizhny Novgorod, Russia

(Received March 12, 1991; revised October 1, 1991; accepted October 15, 1991.)

In this paper the radiation of ring electric and magnetic currents in magnetoactive plasma in whistler frequency range is investigated. Particular attention is concentrated on the problem of variation of the radiated power distribution over the spectrum of excited waves with variation of the radiator parameters. Attention is also given to the radiation pattern structure near the resonance cone and peculiar caustic directions corresponding to the conic refraction and the Storey angle. It is shown that for ring electric current the consideration of the inhomogeneity in its distribution along the ring can be of importance and can lead, in particular, to a large increase of the relative contribution of the quasi-electrostatic whistler mode waves to the total radiation power. Detailed numerical calculations have been carried out for the conditions close to ionospheric ones.

1. INTRODUCTION

A rather large number of papers [Bell and Wang, 1971; Bellyustin, 1978; Golubyatnikov et al., 1988; Wang and Bell, 1969, 1972a, b] is devoted to the problem of radiation of electromagnetic waves by the given electric and magnetic currents in magnetoactive plasma in the whistler frequency range:

$$\omega_{LH} \ll \omega < \omega_H < \omega_p. \quad (1)$$

where ω_p is the electron plasma frequency, ω_H is the gyrofrequency of electrons, ω_{LH} is the lower hybrid frequency. As a rule, linear and ring radiators of the dipole type are considered with the simplest current distributions: triangular and rectangular for linear radiators and homogeneous for ring radiators. Here the problem of radiated energy distribution between proper whistler waves and quasi-electrostatic whistler mode waves induces an increased interest. For separate radiators, in particular a homogeneous ring current, it is shown [Golubyatnikov et al., 1988] that in the case of sufficiently small electric dimensions, $k_0 l v^{1/2} \ll 1$ ($v = \omega_p^2/\omega^2$, $k_0 = (\epsilon_0 \mu_0)^{1/2} \omega$ is the wave number in free space, and l is the effective radiator dimension) practically all the radiated power goes into quasi-

electrostatic whistler mode waves. In the case of comparatively large electric dimensions $k_0 l g^{1/2} \gg 1$ ($g = v/u^{1/2}$, $u = \omega_H^2/\omega^2$) the main part of the power radiates into the proper whistler waves corresponding to the well-known dispersion relation $\eta^2 = g/\cos \vartheta$ [Storey, 1953]. Here η is the refractive index of the whistler wave and ϑ is the angle between the wave vector and the ambient magnetic field. In all intermediate and extensive regions of the radiator parameter values, $k_0 l v^{1/2} \geq 1$ and $k_0 l g^{1/2} \leq 1$, an essential role in radiation field formation is played, evidently by a residual part of the spatial spectrum, covering very interesting and important physical phenomena connected with near-caustic focusing of the electromagnetic field and corresponding to the conic refraction (axial caustic) and the Storey angle (simple caustic) [Arbel and Felsen, 1963; Brodskij et al., 1969, 1972; Al'pert et al., 1983; Lukin et al., 1988]. The presence of peculiar caustic directions, as well as plasma resonance, leads, in turn, to certain difficulties in a correct description of the radiation pattern of the given source [Wang and Bell, 1972b]. The present paper concentrates attention on these problems which, in our opinion, have not received sufficiently clear descriptions and distinct physical interpretations.

We restrict ourselves to the consideration of only ring electric and magnetic currents, raising increased interest particularly in connection with the

Copyright 1992 by the American Geophysical Union.

Paper number 91RS02919.
0048-6604/92/91RS-0291\$08.00

subsequent perspective nonlinear generalizations [Vdovichenko *et al.*, 1986; Markov, 1988]. The corresponding magnetic currents detect a number of principal differences from the similar electric currents. This fact has not been indicated in periodicals, so the case of the magnetic current needs separate discussion. This is more reasonable that definite possibilities of the radiation increase are associated with magnetic currents in some frequency and spatial intervals.

Note that the consideration of the current inhomogeneity in ring radiators is necessary (especially in the case of large electric dimensions) and leads to an essential but not always desired, increase in the energy portion radiated into quasi-electrostatic whistler mode waves. The ordinary homogeneous distribution in the frequency range (1) is a certain degenerate case. However, since such a distribution is used in the majority of papers [Bell and Wang, 1971; Golubyatnikov *et al.*, 1988, Wang and Bell, 1972a], as well as for the possibility of the subsequent comparison, it will be the beginning of our treatment. Again, the uniform distribution sometimes appears to be simply practically preferable, thus it is important to set conditions under which the distribution inhomogeneity of the corresponding ring current can be ignored.

2. BASIC FORMULATION

As it is known, in the given frequency interval (1) the dielectric permittivity tensor of a cold collisionless magnetized ($B_0 = B_0 z^0$) plasma is

$$\hat{\epsilon} = \begin{pmatrix} \epsilon & -ig & 0 \\ ig & \epsilon & 0 \\ 0 & 0 & \eta \end{pmatrix}, \quad (2)$$

where $\epsilon = v/u$, $g = v/u^{1/2}$, $\eta = -v$. Here only the extraordinary wave (the whistler mode) is the propagating one. Its refractive index surface is described by the expression

$$n_{\parallel}^2(n_{\perp}) = \epsilon - \frac{n_{\perp}^2}{2} \left(1 + \frac{1}{\eta} \right) + \sqrt{\frac{n_{\perp}^4}{4} \left(1 - \frac{\epsilon}{\eta} \right)^2 - n_{\perp}^2 \frac{g^2}{\eta} + g^2} \quad (3)$$

and shown in Figure 1. In this figure, symbols \parallel and \perp mark the directions along and transverse the ambient magnetic field and n_{\parallel} , n_{\perp} are the normalized components of the wave vector: $n_{\parallel} = k_{\parallel}/k_0$, $n_{\perp} = k_{\perp}/k_0$, the value $n_{\perp S}$ (see Figure 1) corresponds

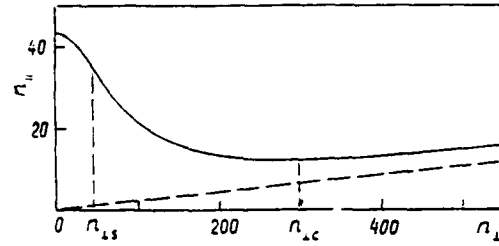


Fig. 1. The whistler mode refractive index surface $n_{\parallel}(n_{\perp})$: $\omega = 1.88 \times 10^5 \text{ s}^{-1}$, $\omega_H = 8.78 \times 10^6 \text{ s}^{-1}$ ($B_0 = 0.5\text{G}$), $\omega_P = 5.64 \times 10^7 \text{ s}^{-1}$ ($N_e = 10^6 \text{ cm}^{-3}$).

to the Storey angle and can be found from the condition $n_{\parallel}'(n_{\perp S}) = 0$, the value $n_{\perp C}$ corresponds to the conic refraction and can be found from the condition $n_{\parallel}'(n_{\perp C}) = 0$ for $n_{\perp C} \neq 0$ (dashes denote differentiation over the argument).

The presentation of the total power P_{Σ} radiated into whistler mode waves by the given harmonical in time ($e^{-i\omega t}$) electric $j^e(\mathbf{r})$ and magnetic $j^m(\mathbf{r})$ currents is as follows:

$$P_{\Sigma} = P_{\Sigma}^e + P_{\Sigma}^m = -\frac{1}{2} \text{Re} \int (j^e, \mathbf{E}) d^3r - \frac{1}{2\mu_0} \text{Re} \int (j^m, \mathbf{B}) d^3r. \quad (4)$$

This formula with the help of simple but expansive transformations is rewritten in the following spectral form convenient for the calculations:

$$P_{\Sigma}^{e,m} = -\frac{k_0^2 z_0}{16\pi^2 \eta} \int \frac{(n^2 - \epsilon)(n_{\perp}^2 - \eta) |F^{e,m}(\mathbf{n})|^2 d^2 n_{\perp}}{n_{\parallel} n_{\perp}^2 [n_{\perp}^4 (1 - \epsilon/\eta)^2 - 4n_{\perp}^2 g^2/\eta + 4g^2]^{1/2}}, \quad (5)$$

$$F^e(\mathbf{n}) = j_x^e(\mathbf{n}) \left(n_x + \frac{ign_y}{n^2 - \epsilon} \right) + j_y^e(\mathbf{n}) \left(n_y - \frac{ign_x}{n^2 - \epsilon} \right) + j_z^e(\mathbf{n}) \frac{n_{\parallel} n_{\perp}^2}{n_{\perp}^2 - \eta}, \quad z_0 = \sqrt{\mu_0/\epsilon_0},$$

$$F^m(\mathbf{n}) = z_0^{-1} \left[j_x^m(\mathbf{n}) n_{\parallel} \left(\frac{\eta n_y}{n_{\perp}^2 - \eta} + \frac{ign_x}{n^2 - \epsilon} \right) + j_y^m(\mathbf{n}) n_{\parallel} \left(\frac{ign_y}{n^2 - \epsilon} - \frac{\eta n_x}{n_{\perp}^2 - \eta} \right) - j_z^m(\mathbf{n}) \frac{ign_{\perp}^2}{n^2 - \epsilon} \right].$$

$$\mathbf{n} = \mathbf{n}_\perp + n_\parallel (\mathbf{n}_\perp \times \mathbf{z}^0).$$

$$j^{e,m}(\mathbf{n}) = \int j^{e,m}(\mathbf{r}) \exp(-ik_0 \mathbf{n} \mathbf{r}) d^3r.$$

To obtain any concrete results, it is necessary to specify the current distribution.

3. HOMOGENEOUS RING ELECTRIC CURRENT

We consider first a circular electric current with uniform symmetrical distribution along the ring:

$$j^e(\mathbf{r}) = \varphi^0 I_0^e \delta(\rho - a) \delta(z) \quad (6)$$

(ρ , φ , z are the cylindrical coordinates). Equation (5) then is transformed into the form:

$$P_\Sigma^e = I_0^{e2} z_0 \frac{\pi}{2} (k_0 a)^2 \frac{g^2}{|\eta|} \int_0^\infty \frac{n_\perp^2 - \eta n_\parallel}{n^2 - \varepsilon} \frac{n_\perp}{n} \frac{J_1^2(k_0 a n_\perp) dn_\perp}{[n_\perp^4 (1 - \varepsilon/\eta)^2 - 4n_\perp^2 g^2/\eta + 4g^2]^{1/2}} \quad (7)$$

where $J_n(\xi)$ is the Bessel function. In the general case the value of the integral incoming into (7) can be obtained only numerically. Some analytical considerations precede the results of the corresponding numerical calculations.

Further, we shall distinguish the following ranges of the integration variable n_\perp :

$$\text{Range I} \quad n_{\perp C} \leq n_\perp < \infty,$$

$$\text{Range II} \quad 0 \leq n_\perp \leq n_{\perp S},$$

$$\text{Range III} \quad n_{\perp S} \leq n_\perp \leq n_{\perp C}.$$

The waves of the range I are the so-called quasi-electrostatic whistler mode waves. The refractive index surface given by the expression (3) is presented approximately for these waves by the formula $\varepsilon n_\perp^2 + \eta n_\parallel^2 = 0$. For whistler mode waves of range II, expression (3) is approximated by the Storey formula $n^2 = g/\cos \vartheta$. The range III corresponds to the "intermediate" region of values of the integration variable n_\perp . We note that for whistler mode waves of this range the approximate relations mentioned above are not valid in the limit of $n_\perp \rightarrow n_{\perp C}$.

First, we obtain the power P_Q^e radiated into quasi-electrostatic whistler mode waves. For this

we take $n_{\perp C} \approx |\eta|^{1/2}$ as a lower integration limit in (7) and regard that for $n_\perp \geq |\eta|^{1/2}$ $n_\parallel \approx (n_\perp^2 - \eta)/n n_\perp^{1/2}$, $n^2 \gg g \gg \varepsilon$. As a result, we derive

$$P_Q^e \approx \frac{2}{3} I_0^{e2} z_0 (k_0 a)^3 g [1 - G_Q(\xi_C)] \quad (8)$$

where

$$G_Q(\xi) = \frac{\pi}{2\xi} \left\{ \left[\xi J_0(\xi) - \frac{1}{2} J_1^2(\xi) \right]^2 + \left(\xi^2 - \frac{3}{4} \right) J_1^2(\xi) \right\}.$$

$\xi_C = k_0 a n_{\perp C} \approx k_0 a |\eta|^{1/2}$. From here, for the radiator of small electric dimensions we come to the well-known formula [Bell and Wang, 1971; Golub-yatnikov et al., 1988]:

$$P_Q^e \approx \frac{2}{3} I_0^{e2} z_0 (k_0 a)^3 g \left[1 - \frac{3\pi}{16} k_0 a |\eta|^{1/2} \right], \quad (9)$$

$$k_0 a |\eta|^{1/2} \ll 1.$$

Second, we calculate a part P_W^e of the power P_Σ^e which is radiated into whistler mode waves corresponding to the range $0 \leq n_\perp \leq n_{\perp S}$. For that we take $n_\perp = n_{\perp S}$ as the upper integration limit in (7). The power P_W^e will be equal to

$$P_W^e = I_0^{e2} z_0 \frac{\pi}{8} g^{-1/2} G_W(\xi_S) \quad (10)$$

where

$$G_W(\xi_S) = \xi_S^2 [J_0^2(\xi_S) + J_1^2(\xi_S)] - 2\xi_S J_0(\xi_S) J_1(\xi_S),$$

$\xi_S = k_0 a n_{\perp S} \approx (2/3)^{1/2} k_0 a g^{1/2}$. In the limiting cases $k_0 a g^{1/2} \ll 1$ and $k_0 a g^{1/2} \gg 1$ we come to formulas

$$P_W^e \approx I_0^{e2} z_0 \frac{\pi}{48} (k_0 a)^4 g^{3/2} \alpha^e, \quad (11)$$

$$P_W^e \approx I_0^{e2} z_0 \frac{1}{4} g^{-1/2} \xi_S \left(1 + \frac{\cos 2\xi_S}{2\xi_S} \right) \quad (12)$$

(α^e is the numerical coefficient of the order of unity).

Third, we estimate the residual part P_f^e of the total radiation power P_Σ^e , which corresponds to an intermediate region of integration variable values n_\perp in (7): $n_{\perp S} \leq n_\perp \leq n_{\perp C}$. Here we failed to obtain general (even such complicated as (8) and (10)) expressions, and thus we have to be satisfied with a

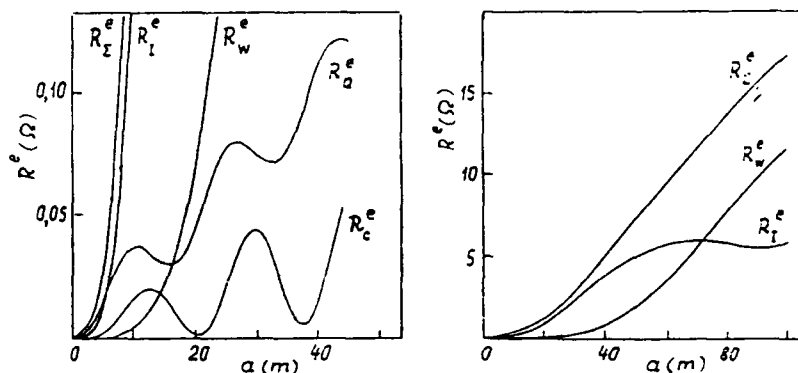


Fig. 2. Total and partial radiation resistances of the uniform ring electric current: $\omega = 1.88 \times 10^5 \text{ s}^{-1}$, $\omega_H = 8.78 \times 10^6 \text{ s}^{-1}$ ($B_0 = 0.5 \text{ G}$), $\omega_p = 5.64 \times 10^7 \text{ s}^{-1}$ ($N_e = 10^6 \text{ cm}^{-3}$).

formula conforming to the limiting case $k_0 a |\eta|^{1/2} \ll 1$:

$$P_f^e \approx I_0^2 z_0 \frac{\pi}{4(2^{3/2})} (k_0 a)^4 g^2 \varepsilon^{-1/2}. \quad (13)$$

To obtain expressions (10)–(13), we used the well-known approximative relations $n^2 \approx g/\cos \vartheta$ ($\cos \vartheta = n_{\parallel}/n$) and $n_{\parallel}(n_{\perp S})/n(n_{\perp S}) \approx 3^{-1/2}$, which are valid in the whistler frequency range (1) practically everywhere but a narrow region of angles $\vartheta \sim \pi/2$. In the framework of the given approximation the coefficient in (11) is $\alpha^e = 0.77$. Note that the chosen partial powers, P_W^e , P_f^e , P_Q^e , overlap completely all the spatial spectrum of excited quasi-plane waves. Thus, due to energetic orthogonality, we have $P_{\Sigma}^e = P_W^e + P_f^e + P_Q^e$.

Last, we evaluate the value P_C^e of power radiated into so-called waves of conic refraction which conventionally can be isolated as $n_{\perp C} - n_{\perp S}/2 \leq n_{\perp C} + n_{\perp S}/2$. It is not difficult to obtain

$$P_C^e \approx I_0^2 z_0 \frac{\pi}{4} (k_0 a)^2 \frac{g}{|\eta|^{1/2}} \sum_{k=0}^{\infty} \left[J_k^2 \left(\xi_C - \frac{1}{2} \xi_S \right) - J_k^2 \left(\xi_C + \frac{1}{2} \xi_S \right) \right] \quad (14)$$

(values ξ_C , ξ_S are defined in (8) and (10)). In the case when $k_0 a g^{1/2} \ll 1$ for $k_0 a n_{\perp C} \neq \nu_m^{(1)}$ ($\nu_m^{(1)}$ are zeros of Bessel function $J_1(\xi)$), (14) is simplified:

$$P_C^e \approx I_0^2 z_0 \frac{\pi}{2} (k_0 a)^2 \frac{g^{3/2}}{|\eta|} J_1^2(k_0 a |\eta|^{1/2}). \quad (15)$$

Thus, for the considered ring radiator of small electric dimensions ($k_0 a |\eta|^{1/2} \ll 1$) we have

$$P_W^e \ll P_C^e, \quad P_W^e \ll P_f^e \ll P_Q^e \sim P_{\Sigma}^e \quad (16)$$

and practically all the power radiates into quasi-electrostatic whistler mode waves, while for the radiator of large electric dimensions ($k_0 a g^{1/2} \gg 1$),

$$P_Q^e \ll P_W^e \sim P_{\Sigma}^e \quad (17)$$

and almost all the power radiates into whistler mode waves of the range $0 \leq n_{\perp} \leq n_{\perp S}$. More accurate and concrete statements can be made on the basis of numerical results which we shall discuss below.

The corresponding results are given in Figure 2 in the form of plots of complete, $R_{\Sigma}^e = 2P_{\Sigma}^e/I_0^2$, and partial, $R_{W,I,Q}^e = 2P_{W,I,Q}^e/I_0^2$, radiation resistances as the functions of the radiator radius for the given values $\omega = 1.88 \times 10^5 \text{ s}^{-1}$, $\omega_p = 5.64 \times 10^7 \text{ s}^{-1}$ (the electron density $N_e = 10^6 \text{ cm}^{-3}$), $\omega_H = 8.78 \times 10^6 \text{ s}^{-1}$ ($B_0 = 0.5 \text{ G}$), which are adequate to the conditions of the Earth's ionosphere (note also the values $|\eta|^{1/2} = 2.2.1$, $g^{1/2} = 43.8$). The partial radiation resistances correspond here as before to the following intervals of values n_{\perp} : R_W^e to the interval $(0, n_{\perp S})$, R_f^e to the interval $(n_{\perp S}, n_{\perp C})$, R_Q^e to the interval $(n_{\perp C}, \infty)$. It is evident that $R_{\Sigma}^e = R_W^e + R_f^e + R_Q^e$. Besides, Figure 2 shows the dependence of the radiation resistance into waves of conic refraction $R_C^e = 2P_C^e/I_0^2$ on the radius a .

According to the data presented in the case of sufficiently small values of a , the main portion of the power radiates into quasi-electrostatic waves. With an increase of a the role of the intermediate

spectral section is gradually increased, so that for $a = a_1 = 5.5$ m the radiation resistance values into quasi-electrostatic and intermediate regions are equaled $R_Q^E(a_1) = R_I^E(a_1)$. Note that in a rather extensive interval of values, a , up to $a = a_2 = 71$ m, where the radiation resistance values into the intermediate region $n_{\perp S} \leq n_{\perp} \leq n_{\perp C}$ and into the region $0 \leq n_{\perp} \leq n_{\perp S}$ are equaled $R_I^E(a_2) = R_W^E(a_2)$, just the intermediate region dominates. Only at rather large values of a ($a > a_2$) the main portion of the radiated power goes into the whistler mode waves of the range II: $0 \leq n_{\perp} \leq n_{\perp S}$. We notice that R_Q^E has a pronounced oscillation character. This fact tells essentially the value of the radiation pattern along the ambient magnetic field.

4. INHOMOGENEOUS RING ELECTRIC CURRENT

Let us consider inhomogeneous ring electric current, the density of which is given in the form:

$$j^e(r) = \varphi^0 \left\{ I_0^e + i_0^e \begin{pmatrix} \cos m\varphi \\ \sin m\varphi \end{pmatrix} \right\} \delta(\rho - a) \frac{B}{\pi(B^2 + z^2)}, \quad (18)$$

$$m = 1, 2, \dots$$

The necessity of "spreading" of the current distribution over the coordinate z is associated with the fact that in contrast to the uniform electric current (6) the total radiated power here, as we shall see, tends to infinity in the case $B \rightarrow 0$. For a comparison we consider the same particular cases which have been discussed above applicable to a uniform ring current (6). The results given here refer only to the simplest inhomogeneous distribution with $m = 1$ and the simplest limiting cases.

The power $P_{Q,1}^E$ radiating into quasi-electrostatic whistler mode waves in the case $k_0 B \epsilon^{1/2} \ll k_0 a |\eta|^{1/2} \ll 1$ is described by the expression

$$P_{Q,1}^E \approx I_0^e z_0 \frac{2}{3} (k_0 a)^3 g + i_0^e z_0 \frac{\pi}{16} (k_0 a)^2 \frac{g}{|\eta|^{1/2}} + i_0^e z_0 \frac{1}{4 k_0 a \epsilon^{1/2} |\eta|^{1/2}} \left[\ln \left(\frac{4a |\eta|^{1/2}}{B \epsilon^{1/2}} \right) - 2 \right]. \quad (19)$$

The last term in (19) is responsible for the occurrence of the effective transverse (relative to B_0) electric dipole moment making for the case of small radius a , the main contribution into the value $P_{Q,1}^E$. So, the power radiated by the current (18) into

quasi-electrostatic whistler mode waves, and hence the total radiated power at $B \rightarrow 0$ (in the case of transition in (18) to δ function over z) are logarithmically diverged, unlike the case of uniform ring electric current (6) where these powers remain finite and can be described by the formula (8). The corresponding differences are obviously manifested in the character of excited spatial spectrum: the distribution (18) at $B \ll a$, $k_0 a |\eta|^{1/2} \ll 1$ in the case $i_0^e = 0$ (uniform distribution), most effectively excites quasi-electrostatic whistler mode waves with typical transverse scale $2\pi/k_0 n_{\perp} \sim a$. In the case of inhomogeneous distribution ($i_0^e \neq 0$), quasi-electrostatic whistler mode waves are effectively excited and with smaller scales up to $2\pi/k_0 n_{\perp} \sim B \ll a$, that finally leads to divergence at the tendency B to zero. Hence uniform azimuthal symmetric distribution (6) is a somewhat degenerate case, since for $B \rightarrow 0$, any (even very small) breaking of homogeneity cardinally changes the spatial spectrum of excited quasi-electrostatic whistler mode waves.

The power $P_{W,1}^E$ radiated into the whistler mode waves of the range $0 \leq n_{\perp} \leq n_{\perp S}$ at $k_0 a g^{1/2} \ll 1$, $B \ll a$, and $k_0 a g^{1/2} \gg 1$, $k_0 B g^{1/2} \ll 1$ has the form

$$P_{W,1}^E = I_0^e z_0 \frac{\pi}{48} (k_0 a)^4 g^{3/2} \alpha^e + i_0^e z_0 \frac{\pi}{16 \sqrt{3}} (k_0 a)^2 g^{1/2} \beta^e. \quad (20)$$

$$P_{I,1}^E \approx I_0^e z_0 \frac{1}{4 g^{1/2}} \xi_S \left(1 + \frac{\cos 2\xi_S}{2\xi_S} \right) + i_0^e z_0 \frac{\xi_S}{8 g^{1/2}} \quad (21)$$

(overestimations $\alpha^e = 0.77$, $\beta^e = 1.15$).

The power $P_{I,1}^E$ radiated into the intermediate region $n_{\perp S} \leq 0 \leq n_{\perp C}$, at $k_0 B g^{1/2} \ll 1$, $k_0 a |\eta|^{1/2} \ll 1$ accepts the form

$$P_{I,1}^E \approx I_0^e z_0 \frac{\pi}{8 \sqrt{2}} (k_0 a)^4 \frac{g^2}{\epsilon^{1/2}} + i_0^e z_0 \frac{\pi}{48} (k_0 a)^2 \frac{\eta^{3/2}}{g}. \quad (22)$$

At last, the power $P_{C,i}^E$ radiated into waves of conic refraction in the case $k_0 |\eta|^{1/2} \ll 1$, $k_0 B \epsilon^{1/2} \ll 1$ is equaled to

$$P_{C,1}^E \approx I_0^e z_0 \frac{\pi}{8} (k_0 a)^4 g^{3/2} + i_0^e z_0 \frac{\pi}{16} (k_0 a)^2 \frac{|\eta|}{g^{1/2}}. \quad (23)$$

So for a ring radiator of small electric dimensions $k_0 B g^{1/2} \ll k_0 a |\eta|^{1/2} \ll 1$ with inhomogeneous distribution of the current (18), almost all the radiating

TABLE 1. Values a_1 and a_2 for inhomogeneous ring electric current, $\omega = 1.88 \times 10^5 \text{ s}^{-1}$, $\omega_H = 8.78 \times 10^6 \text{ s}^{-1}$, $\omega_p = 5.64 \times 10^7 \text{ s}^{-1}$, $B = 10^{-2} \text{ m}$

| i_0^e/I_0^e | a_1, m | a_2, m |
|----------------------|-----------------|-----------------|
| 0 | 5.5 | 71.0 |
| 0.01 | 7.2 | 71.8 |
| 0.05 | 12.2 | 72.0 |
| 0.10 | 16.7 | 72.1 |
| 0.50 | 48.0 | 75.0 |
| $\infty (I_0^e = 0)$ | 177.0 | 126.0 |

power goes into quasi-electrostatic whistler mode waves:

$$P_{W,1}^e < P_{C,1}^e, \quad P_{W,1}^e \ll P_{I,1}^e \ll P_{Q,1}^e \sim P_{\Sigma,1}^e. \quad (24)$$

With an increase of the loop radius a the radiated power is redistributed over the spectrum of excited waves, similar to the above considered case of uniform ring electric current (6). An essential difference is, however, in the fact that the values a_1 and a_2 for which the corresponding partial powers are equaled to $P_{Q,1}^e(a_1) = P_{I,1}^e(a_1)$, $P_{I,1}^e(a_2) = P_{W,1}^e(a_2)$, in the case of inhomogeneous current ($i_0^e \neq 0$) are shifted to the side of larger values. It is obviously illustrated by Table 1 for the simplest distribution with $m = 1$ in (18) and earlier chosen (Figure 2) parameters ω , N_e , B_0 and $B = 10^{-2} \text{ m}$.

The dependence of both total $R_{\Sigma,1}^e = 2P_{\Sigma,1}^e/I_0^{e2}$ and partial $R_{Q,1}^e = 2P_{Q,1}^e/I_0^{e2}$, $R_{I,1}^e = 2P_{I,1}^e/I_0^{e2}$, $R_{W,1}^e = 2P_{W,1}^e/I_0^{e2}$ radiation resistances on the radius a , for one of the values $i_0^e/I_0^e = 0.5$ given in Table 1 are plotted in Figure 3. We notice that the increasing of values $R_{Q,1}^e(a)$ and $R_{\Sigma,1}^e(a)$ with the decreasing of a is typical for the corresponding electric dipole. The increase occurs up to the anomalously small values $a \sim B\epsilon^{1/2}|\eta|^{-1/2}$ which are of no practical interest. For $a \rightarrow 0$, values $R_{Q,1}^e$ and $R_{\Sigma,1}^e$ naturally drop to zero.

We have derived the radiated power $P_{Q,m}^e$ within the cold collisionless plasma approximation. For more realistic models of the ionosphere the electron collisions and warm-plasma effects should be taken into consideration. These two effects will limit the maximum real value n_{\perp} , so that the integral (5) will possess a finite upper limit in n_{\perp} . Calculations show that in a plasma with electron temperatures and electron collision frequencies typical of the ionosphere the radiation resistance $R_{Q,1}^e$ of the inhomogeneous current (18) with dimensions $a \geq 1 \text{ m}$, $B \geq 10^{-2} \text{ m}$ is of the same order as the value $R_{Q,1}^e$

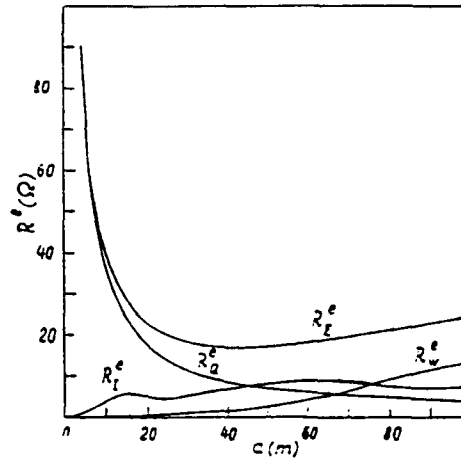


Fig. 3. Total and partial radiation resistances of the inhomogeneous ring electric current: $B = 10^{-2} \text{ m}$, $i_0^e/I_0^e = 0.5$, $m = 1$, $\omega = 1.88 \times 10^5 \text{ s}^{-1}$, $\omega_H = 8.78 \times 10^6 \text{ s}^{-1}$, $\omega_p = 5.64 \times 10^7 \text{ s}^{-1}$.

obtained for the idealized plasma model. Consequently, the large increase of magnitude of $R_{Q,m}^e$ caused by the current inhomogeneity will occur in the real ionosphere as well.

5. RING MAGNETIC CURRENT

Now we consider uniform ring magnetic current, the density of which is

$$j^m(r) = \varphi_0 I_0^m \delta(\rho - a) \frac{B}{\pi(B^2 + z^2)}. \quad (25)$$

The necessity of the corresponding "spreading" of current is due to the fact that here, unlike the electric current (6), already uniform distribution at $B \rightarrow 0$ leads to the divergence of the total radiated power.

The power P_Q^m taken away by quasi-electrostatic whistler mode waves at $k_0 B \epsilon^{1/2} \ll k_0 a |\eta|^{1/2} \ll 1$ and $k_0 a |\eta|^{1/2} \gg 1$, $k_0 B \epsilon^{1/2} \ll 1$ is given in the form

$$P_Q^m \approx I_0^{m2} z_0^{-1} \frac{1}{2} (k_0 a) g \left[\ln \left(\frac{4a}{B} \frac{|\eta|^{1/2}}{\epsilon^{1/2}} \right) - 2 \right], \quad (26)$$

$$P_Q^m \approx I_0^{m2} z_0^{-1} \frac{1}{2} (k_0 a) g \left[\ln \frac{4}{k_0 B \epsilon^{1/2}} - 2 \right]. \quad (27)$$

These formulas testify to the divergence of the radiated power. As in the previous case, the divergence is of the logarithmic character.

The power P_W^m radiated into whistler mode waves

of the range $0 \leq n_{\perp} \leq n_{\perp S}$ accepts the form $(k_0 B g^{1/2} \ll 1)$:

$$P_W^m = I_0^{m^2} z_0^{-1} \frac{\pi}{48} (k_0 a)^4 g^{5/2} \alpha^m, \quad k_0 a g^{1/2} \ll 1. \quad (28)$$

$$P_W^m = I_0^{m^2} z_0^{-1} \frac{g^{1/2}}{4} \xi_S \left(1 + \frac{\cos 2\xi_S}{2\xi_S} \right), \quad k_0 a g^{1/2} \gg 1. \quad (29)$$

(for the approximations used, $\alpha^m = 1.12$). The power P_I^m going into "intermediate" whistler mode waves at $k_0 B g^{1/2} \ll 1$, $k_0 a |\eta|^{1/2} \ll 1$ is given by the formula

$$P_I^m = I_0^{m^2} z_0^{-1} \frac{\pi}{24} (k_0 a)^4 \varepsilon^{1/2} \eta^2. \quad (30)$$

In the case of large electric dimensions of the source (25) we did not succeed in obtaining the general analytical expression for the value P_I^m ; yet we can show that under the conditions $k_0 a g^{1/2} \gg 1$, $k_0 B g^{1/2} \ll 1$ the inequality takes place:

$$P_I^m \leq I_0^{m^2} z_0^{-1} \frac{1}{2} k_0 a g \ln \frac{|\eta|^{1/2}}{g^{1/2}}. \quad (31)$$

From the formulas obtained it follows that with the increase of the radiator radius a , in spite of an increase of the relative portion of the energy radiated into the ranges II and III, a larger part of the radiated power goes as before into quasi-electrostatic whistler mode waves. This conclusion is confirmed by the detailed numerical calculations. So the character of the radiated power redistribution as a function of the radiator dimension a for the magnetic current (25) differs markedly from the case of ring electric current. This is due to differences in the structure of electric and magnetic components of the field excited by ring electric and magnetic currents and can be followed on the basis of the general expression (4).

Taking into account the magnetic current inhomogeneity along the ring is not as important here as it is for the electric current and does not lead to a considerable increase of the total radiating power nor the relative portion of the energy going into quasi-electrostatic whistler mode waves. At the same time, the current inhomogeneity under certain conditions can assist in relatively increasing the power radiated into the ranges II and III. We note that electron collisions and warm-plasma effects will not change all these conclusions. The analytical

and numerical results are not given here for the sake of brevity.

6. RADIATION PATTERN OF RING CURRENTS

We shall define the radiation pattern for the corresponding current as

$$D(\vartheta, \varphi) = \lim_{\Delta\Omega \rightarrow 0} \frac{1}{\Delta\Omega} \lim_{r \rightarrow \infty} \int_{\Delta\Omega} S_r(r, \vartheta, \varphi) r^2 d\Omega \quad (32)$$

where r, ϑ, φ are the spherical coordinates, S_r is the radial component for the Poynting vector S , $d\Omega = \sin \vartheta d\vartheta d\varphi$ the element of the solid angle.

We give the resulting expression for the radiation pattern of the simplest uniform ring electric current (6) obtained with the use of the stationary phase method:

$$\begin{aligned} D(\vartheta, \varphi) &= \sum_i D_i(\vartheta, \varphi) \\ &= I_0^2 z_0^{-1} \frac{1}{8} (k_0 a)^2 \sum_i \sin^{-1} \vartheta P_i(\gamma_i, \varphi) \\ &\quad \cdot \frac{n(\gamma_i) \sin \gamma_i}{[(\varepsilon^2 - g^2 - \varepsilon \eta)^2 \sin^4 \gamma_i - 4g^2 \eta^2 \cos^2 \gamma_i]^{1/2}} \\ &\quad \cdot \left| \frac{n^2(\gamma_i) + n'^2(\gamma_i)}{n^2(\gamma_i) + 2n'^2(\gamma_i) - n(\gamma_i)n''(\gamma_i)} \right| \end{aligned} \quad (33)$$

where

$$P(\gamma, \varphi) = g^2 \frac{n^2(\gamma) \sin^2 \gamma - \eta}{n^2(\gamma) - \varepsilon} J_1^2[k_0 a n(\gamma) \sin \gamma],$$

$$n(\gamma) = n_2(\gamma)/n_1(\gamma), \quad n_1(\gamma) = (\eta \cos^2 \gamma - \varepsilon \sin^2 \gamma)^{1/2},$$

$$n_2(\gamma) = 2^{-1/2} \{ (\varepsilon \eta + g^2 - \varepsilon^2) \sin^2 \gamma - 2\varepsilon \eta - [(\varepsilon \eta + g^2 - \varepsilon^2)^2 \sin^4 \gamma + 4g^2 \eta^2 \cos^2 \gamma]^{1/2} \}^{1/2},$$

$\gamma = \arctan [n_{\perp}/n_{\parallel}(n_{\perp})]$ in the polar angle in the space of wave vectors. Summation in (33) covers all solutions $\gamma_i = \gamma_i(\vartheta)$ of the equation

$$\vartheta = \gamma + \arctan [n'(\gamma)/n(\gamma)], \quad \gamma \in [0, \gamma_C], \quad i = 1, 2,$$

$$\vartheta = -\gamma + \arctan [n'(\gamma)/n(\gamma)], \quad \gamma \in [\gamma_C, \gamma_r], \quad i = 3,$$

where $\gamma_C = \arctan (n_{\perp C}/n_{\parallel}(n_{\perp C}))$, $\gamma_r = \arctan [|\eta|^{1/2}/\varepsilon^{1/2}]$. In reality, in formula (33), because of

energetic orthogonality, we sum the energy fluxes responsible for separate rays transforming the energy in the given direction (ϑ, φ) . Note that the expression for the radiation pattern can be derived also by transforming the spectral representation of the total radiated power (5) in the wave normal space to the integral over the spherical coordinates ϑ and φ in the observation space [Wang and Bell, 1972b].

As seen from Figure 1, in the angle interval $0 < \vartheta < \vartheta_r$ ($\vartheta_r = \pi/2 - \gamma_r$), three rays correspond to each of the chosen directions (ϑ, φ) , in the angle interval $\vartheta_r < \vartheta < \vartheta_S$, two rays (here ϑ_S is the Storey angle, ϑ_r is the angle at which the resonance cone occurs). At the boundary of the propagation region $\vartheta \rightarrow \vartheta_S$ these rays coincide forming a simple caustic surface. In the direction of the external magnetic field $\vartheta = 0$, one ordinary ray corresponding to the angle $\gamma = 0$ occurs. Besides, in this direction an infinite number of rays corresponding to the conic refraction (i.e., to the angle $\gamma = \gamma_C$) coincides and forms the axial caustic.

Expression (33) corresponds in reality to the standard geometrical optics approximation and, generally speaking, are not applicable near the caustic directions. However, since in fulfilling the limit transition $r \rightarrow \infty$, the angular width of the near-caustic region tends to zero, we can obtain the correct expressions for the radiation pattern near the caustic directions using formula (33). This is confirmed by the results of more detailed investigations. As a result, for the radiation pattern near the axial caustic $\vartheta \rightarrow 0$ and simple caustic $\vartheta \rightarrow \vartheta_S$ we have

$$\lim_{\vartheta \rightarrow 0} D(\vartheta, \varphi) = D_C / \sin \vartheta \quad (34)$$

where

$$D_C = I_0^2 \epsilon_0 \frac{1}{4} (k_0 a)^2 g^2 \frac{n^2(\gamma_C) \sin \gamma_C - \eta}{n^2(\gamma_C) - \epsilon} n(\gamma_C) \sin \gamma_C \cdot \frac{J_1^2(k_0 a n(\gamma_C) \sin \gamma_C)}{[(\epsilon^2 - g^2 - \epsilon \eta)^2 \sin^4 \gamma_C + 4g^2 \eta^2 \cos^2 \gamma_C]^{1/2}} \cdot \left| \frac{n^2(\gamma_C) + n'^2(\gamma_C)}{n^2(\gamma_C) + 2n'(\gamma_C) - n(\gamma_C)n''(\gamma_C)} \right| \cdot \lim_{\vartheta \rightarrow \vartheta_S} D(\vartheta, \varphi) = D_S(\vartheta_S - \vartheta)^{-1/2} \quad (35)$$

where

$$D_S = I_0^2 \epsilon_0 \frac{1}{4} (k_0 a)^2 g^2 \frac{n^2(\gamma_S) \sin^2 \gamma_S - \eta}{n^2(\gamma_S) - \epsilon} n(\gamma_S) \sin \gamma_S \cdot \frac{J_1^2(k_0 a n(\gamma_S) \sin \gamma_S)}{[2n''(n_{\perp S})]^{1/2} [(\epsilon^2 - g^2 - \epsilon \eta)^2 \sin^4 \gamma_S + 4g^2 \eta^2 \cos^2 \gamma_S]^{1/2} i g \gamma_S} \cdot \frac{n^2(\gamma_S) + n'^2(\gamma_S)}{[n'(\gamma_S) \sin \gamma_S + n(\gamma_S) \cos \gamma_S]^2}.$$

From these expressions it follows that $D(\vartheta, \varphi)$ is diverged for $\vartheta \rightarrow 0$ as ϑ^{-1} and for $\vartheta \rightarrow \vartheta_S$ as $(\vartheta_S - \vartheta)^{-1/2}$; here these divergences are integrated, so that the power radiated in any finite solid angle $\Delta\Omega$ is the finite value.

In the region of angles $\vartheta \rightarrow \vartheta_r$, the radiation pattern is correctly described by geometrical optics expression (33). It is easy to see that the peculiarity of the radiation pattern corresponding to the resonance cone $\vartheta = \vartheta_r$ is the integrated one. Really, making the transition $\vartheta \rightarrow \vartheta_r - 0$ in (33), we note that the contribution to the radiation pattern from rays satisfying the resonance cone is finite and has the form

$$D_r(\vartheta, \varphi) = D_3(\vartheta, \varphi) = I_0^2 \epsilon_0 \frac{1}{8\pi} k_0 a \frac{g^2}{n_2(\gamma_r)} \frac{\sin^2 \gamma_r}{\sin \vartheta_r} \cdot \{1 - \sin [2k_0 a n_2(\gamma_r) \sin \gamma_r n_1(\gamma_r)]\}. \quad (36)$$

From (36) it is seen that the function $D(\vartheta, \varphi)$ near the direction $\vartheta = \vartheta_r$ has oscillations over ϑ , the frequency of which continuously increases up to infinity at $\vartheta \rightarrow \vartheta_r - 0$ [$n_1(\gamma) \rightarrow 0$].

As an example, Figure 4 presents the radiation pattern of the current (6) normalized by I_0^2 (curve 1) with the radius $a = 10$ m and the given values $\omega = 1.88 \times 10^5$ s⁻¹, $\omega_H = 8.78 \times 10^6$ s⁻¹, $\omega_P = 5.64 \times 10^7$ s⁻¹; and the values of auxilliary functions $\bar{D}_C(\vartheta) = \int_0^{2\pi} D(\vartheta, \varphi) \sin \vartheta d\varphi$ (curve 2) and $\bar{D}_S(\vartheta, \varphi) = D(\vartheta, \varphi) \sin \vartheta (\varphi_S - \vartheta)^{1/2}$ (curve 3) normalized by I_0^2 . In the case considered it is evident

$$\bar{D}_C(0) = 2\pi D_C, \quad \bar{D}_S(\vartheta_S, \varphi) = D_S \sin \vartheta_S.$$

For comparison, Figure 5 shows the dependence of the product $S_r r^2$ on the angle ϑ near peculiar directions $\vartheta = 0$, $\vartheta = \vartheta_S$ for different distances r from the radiator up to the observation point. Wang and Bell [1972b] suggested defining the radiation pattern in a peculiar direction through the near-

caustic section of the function $S_r r^2$ (see Figure 5). The angular dependence of the value $S_r r^2$ calculated by this method depends essentially on the distance r and in the near-caustic region does not coincide with the radiated power flux per unit solid angle. Again, the problem occurs in the correct "matching" of the near-caustic section of the radiation pattern with its geometrical optics part. It is more convenient and appropriate to the physical nature of the radiation pattern to use the usual integral definition (32) and auxiliary functions \tilde{D}_C , \tilde{D}_S , which are finite, in the isolation of the character of peculiarity.

7. CONCLUSION

We have shown that in the frequency range (1) when the refractive index surface is opened, the consideration of inhomogeneity in the distribution of the ring electric current is of importance and leads to a large increase in the power radiated into quasi-electrostatic whistler mode waves. If such an increase is undesirable, it is necessary to provide the maximal uniformity of the distribution. Conditions of such "uniformity," i.e., conditions for the limiting permissible inhomogeneity (in our case for the relation i_0^e/I_0^e) which is yet nonessential, are derived from the above formulas and, generally speaking, are rather rigid. For ring magnetic cur-

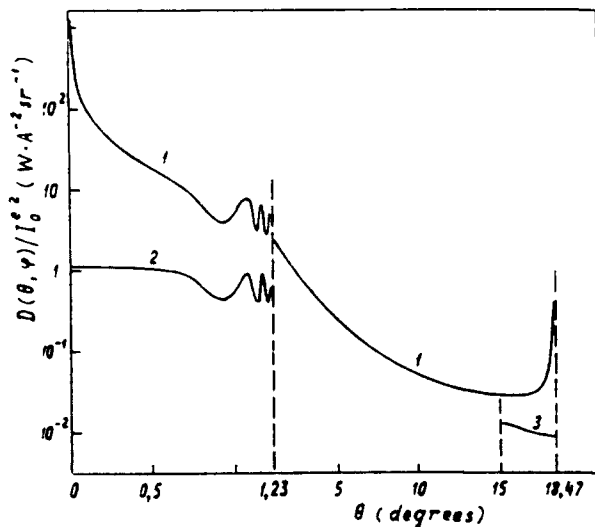


Fig. 4. Radiation pattern of the uniform ring electric current $a = 10$ m. values of ω , ω_H , and ω_p are similar to Figures 1-3.

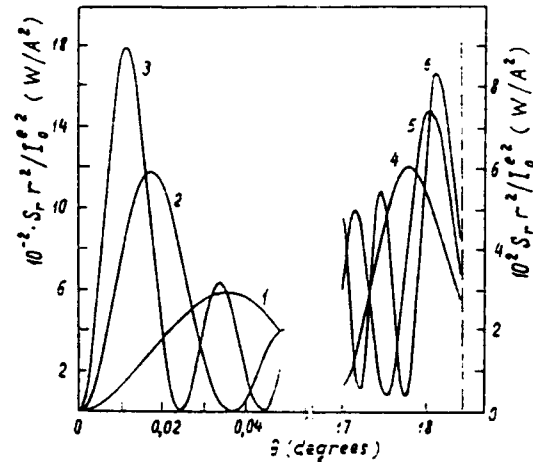


Fig. 5. Angular dependence of the product $S_r r^2$ near peculiar caustic directions: curves 1, 4 - $k_0 r = 10$; curves 2, 5 - $k_0 r = 20$; curves 3, 6 - $k_0 r = 30$; $a = 10$ m. values of ω , ω_H , and ω_p are similar to Figures 1-3.

rents the consideration of the inhomogeneity of the distribution along the ring is not as important as for their electric analogies.

All the given investigations testify to the necessity and importance of the developing the theory of thin antennas in such media which would permit one to find the real current distributions. Again, it is clear that even under linear conditions the action of the antenna on the surrounding plasma can lead to marked variations of radiation characteristics.

Finally, it should be noted that, as well as in all referenced works, we have considered the case of the cold homogeneous plasma with the homogeneous ambient magnetic field. These simplifications have allowed us to concentrate particular attention on the most important physical aspects of the problems discussed.

REFERENCES

- Al'pert, Ya. L., K. G. Budden, B. S. Moiseyev, and G. F. Stott, Electromagnetic radiation from a dipole source in a homogeneous magnetoplasma, *Philos. Trans. R. Soc. London, Ser. A*, 309(11), 503-557, 1983.
- Arbel, E., and L. B. Felsen, Theory of radiation from sources in an anisotropic media, part II, in *Proceedings of the Symposium on Electromagnetic Theory and Antennas*, Part 1, edited by E. C. Jordan, pp. 421-459, Pergamon, New York, 1963.
- Bell, T. F., and T. N. C. Wang, Radiation resistance of a small filamentary loop antenna in a cold multicomponent magnetoplasma, *IEEE Trans. Antennas Propag.*, AP-19(4), 517-522, 1971.

- Bellyustin, N. S., Wave radiation of whistler-range in plasma, *Izv. Vyssh. Uchebn. Zaved. Radiofiz.*, 21(1), 22-35, 1978.
- Brodskij, Yu. Ya., I. G. Kondrat'ev, and M. A. Miller, Electromagnetic beams in anisotropic media, *Alta Freq.*, 38, spec. issue, 89-93, 1969.
- Brodskij, Yu. Ya., I. G. Kondrat'ev, and M. A. Miller, Electromagnetic beams in anisotropic media, II, *Izv. Vyssh. Uchebn. Zaved. Radiofiz.*, 15(4), 592-600, 1972.
- Golubyatnikov, G. Yu., S. V. Egorov, A. V. Kostrov, E. A. Mareev, and Yu. V. Chugunov, Excitation of electrostatic and whistler waves by a magnetic type antenna, *Zh. Eksp. Teor. Fiz.*, 94(4), 124-135, 1988.
- Lukin, D. S., V. B. Presnyakov, and P. P. Savchenko, Calculation of wave fields in the near zone of frame VLF-emitter in the uniform magnetoactive plasma, *Geomagn. Aeron.*, 28(2), 218-221, 1988.
- Markov, G. A., Observation of a resonant autotuning of a magnetic antenna by rf-discharge plasma, *Fiz. Plasmy. Moscow*, 14(9), 1094-1098, 1988.
- Storey, L. R. O., An investigation of whistling atmospherics, *Philos. Trans. R. Soc. London, Ser. A*, 246(908), 113-141, 1953.
- Vdovichenko, I. A., G. A. Markov, V. A. Mironov, and A. M. Sergeev, Ionization self-channeling of whistler waves in plasma, *Pis'ma Zh. Eksp. Teor. Fiz.*, 44(5), 216-219, 1986.
- Wang, T. N. C., and T. F. Bell, Radiation resistance of a short dipole immersed in a cold magnetoionic medium, *Radio Sci.*, 4(2), 167-177, 1969.
- Wang, T. N. C., and T. F. Bell, VLF-ELF input impedance of an arbitrarily oriented loop antenna in a cold collisionless multicomponent magnetoplasma, *IEEE Trans. Antennas Propag.*, AP-20(3), 394-398, 1972a.
- Wang, T. N. C., and T. F. Bell, VLF-ELF radiation patterns of arbitrarily oriented electric and magnetic dipoles in a cold lossless multicomponent magnetoplasma, *J. Geophys. Res.*, 77(7), 1174-1189, 1972b.

I. G. Kondrat'ev and A. V. Kudrin, Department of Radiophysics, University of Nizhnij Novgorod, Nizhnij Novgorod, Russia.
T. M. Zaboronkova, Radiophysical Research Institute, Nizhnij Novgorod, Russia.

The Lorentz polarization term for magnetospheric whistler propagation

Hal J. Strangeways

Department of Electronic and Electrical Engineering, University of Leeds, Leeds, U.K.

(Received February 4, 1991, revised April 24, 1991, accepted May 1, 1991)

If the particles of a dielectric are arranged in a cubic lattice, the effective field acting on a particle is $E + lP/\epsilon_0$, where $l=1/3$. The second term is the Lorentz polarization term. Appleton believed that this term should be omitted from magneto-ionic theory, whereas Hartree included it. Various arguments have been presented to prove Appleton's contention that this term is zero for a plasma. Some proofs have considered the perturbation of the individual orbits of the electrons, but careful consideration of these treatments shows that they would predict a nonzero Lorentz term for some velocity distributions. It has also been argued that $l \neq 0$ would violate causality but it can be shown that this only applies in the zero frequency limit for positive l . Other proofs that have been presented also possess significant limitations, in particular being limited to rather simpler plasma configurations than exist in the magnetosphere and weak field strengths. The theoretical problem for more general conditions is very difficult and thus an appeal to experiment results is considered relevant. It is shown that a very small nonzero Lorentz term would be significant for whistler-mode propagation. The observed lower cutoff frequencies of whistlers put a very low limit ($l \sim 10^{-5}$) on the maximum value of positive l but put no limit on negative l . It is shown however, that $l < 0$ would increase the cutoff frequency for ducted propagation. Thus the upper cutoff frequency is determined for values of l in the range -5×10^{-6} to -2×10^{-3} by ray tracing in a model Gaussian cross-section duct in a realistic model of the magnetospheric plasma corresponding to winter night conditions. Cutoffs (normalized to the equatorial electron gyrofrequency determined by curve fitting to the ray-traced whistler spectra) are found to be in the range 0.49 to 0.55 for a 15% enhancement duct. The range of cutoffs for $l \sim -10^{-4}$ is in much better agreement with experimentally observed whistler cutoffs than the assumption that $l=0$ (no Lorentz term). This provides experimental evidence for a nonzero (negative) Lorentz term for ducted whistler propagation.

1. INTRODUCTION

It was found by Lorentz [1909] that the electric field acting on an individual particle in a dielectric was not simply equal to the applied field but contained an extra term corresponding to an additional electric field resulting from the displacement of the electrons in all the neighboring particles. Lorentz considered passing an electromagnetic wave through a medium containing particles, each of which could perform damped oscillations;

$$m\ddot{x} + \alpha\dot{x} + \beta x = eE' \quad (1)$$

e and m are the charge and mass of an electron, respectively, and E' is the electric field acting on each particle including that produced by the displacement of the electrons in the neighboring particles. In order to calculate E' , Lorentz

divided the medium in two. The medium was considered as a smoothed-out continuum everywhere outside a sphere of radius r centered on the particle under consideration, but inside this sphere the fields of the individual particles were considered. This was because the smoothing out operation was considered reasonable at distances very much greater than the inter-particle distance but not for comparable distances. Thus the radius of the sphere should be large compared with the inter-particle distance but must also be small compared with the wavelength. The electric field acting on a particle is then given by the applied field (E) plus that due to the continuum outside the sphere (E_1) plus that due to the particles inside the sphere E_2 . Thus

$$E' = E + E_1 + E_2 \quad (2)$$

Lorentz showed that the field at the center of the sphere, due to the external continuum alone (E_1), was independent of r and was given by $P/3\epsilon_0$, where P was the volume polarization of the continuum. The field due to the spherical distribution of charge (E_2) was shown to be zero for a cubic

Copyright 1992 by the American Geophysical Union

Paper number 91RS02114.
0048-6604/92/91RS-02114\$08.00

lattice. It is commonly stated [e.g. Ratcliffe, 1959; Budden, 1985; Fejer, 1985] that Lorentz also showed that $E_2 = 0$ would also be true for a random distribution of dipoles. In fact, Lorentz [1909] stated that for isotropic bodies such as glass, fluids, and gases, $E_2 = sP$, where s is a constant which it would be difficult to exactly determine. Other authors have shown this field to be zero for a random distribution of particles [e.g. Chelkowski, 1980]. However, the distribution of electrons (or ions) in a plasma cannot be considered to be completely random since electrostatic repulsive forces will obviously reduce the number of cases where the separation between electrons (or ions) is very small. Thus the separation between electrons will be more strongly peaked near $N^{-1/3}$ than a purely random distribution. In spite of this, for a plasma it is always assumed that $E_2 = 0$. Thus

$$E' = E + P/3\epsilon_0 \quad (3)$$

The extra term is known as the Lorentz polarization term. It is convenient to write it as lP/ϵ_0 . The value $l=0$ would then correspond to noninclusion of the term and $l=1/3$ the case of the cubic lattice or random distribution. Other values of l including negative values will, however, also be considered. We will refer to terms of the form lP/ϵ_0 as Lorentz terms regardless of whether $l=1/3$. When $l<0$, the terminology anti-Lorentz will also be used.

When the index of refraction of the ionosphere was first determined, it was not immediately clear whether the Lorentz term should be included or not. Lassen [1927] and Appleton [1932] considered that this term should be omitted from magneto-ionic theory, whereas Hartree [1929] included it. The situation is clearly different from that of a solid dielectric in that the electrons are freely moving and not bound to any individual atom or molecule. Nevertheless each electron, displaced by an applied field, can be shown to be equivalent to adding a dipole as shown in Figure 1. If the static situation is considered, it would appear that there should be a Lorentz term when the virtual dipoles are short compared with their separation as would generally occur for HF waves in the ionosphere but that the Lorentz term would disappear if the length of the virtual dipoles became comparable to their separation as would be the case at very low frequencies, approaching the $\omega=0$ limit. For dc, the Lorentz term would

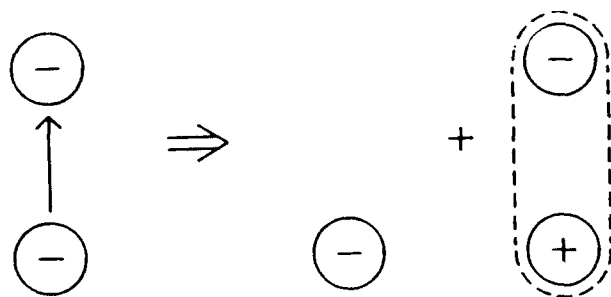


Fig. 1. Electron displacement equivalent to undisplaced electron plus virtual dipole.

need to be zero as otherwise Ohm's law would not be valid [c.f. Ginzburg, 1970]. Thus in the early days there was some considerable debate about the retention or not of the Lorentz term in magneto-ionic theory. Experimental evidence, particularly that from whistlers showed that $l=1/3$ was impossible so that it was generally concluded that l was zero. However, it is pertinent to consider the question of whether l is always identically equal to zero or whether it can take some (very small) residual value (either positive or negative) for some plasma configurations and conditions. Although the question of the Lorentz term has been debated for many years, a consideration of the various arguments employed show that it is not easy to decide this matter solely on theoretical grounds.

2. LIMITATIONS OF PROOFS THAT THE LORENTZ TERM IS ZERO

Although various proofs of the Lorentz term being zero have been proposed they all have some limitations [Strangeways, 1991b]. In general they apply to particular simple plasma configurations such as homogeneous plasma, isotropic or single-valued velocity distribution with ion motion not being properly considered. The only exception to this is the work of Fejer [1985] who argued that a nonzero value of l in the term lP/ϵ_0 would result in causality being violated. Thus his proof, if valid, would rule out even a very small value of l and would apply to a very wide range of plasma configurations. Fejer showed that a nonzero Lorentz term would lead to a pole below the real ω axis in the formula for the permittivity, which would correspond to an exponentially increasing field that would violate causality. However, this pole occurs for $\text{real}(\omega)=0$ so that there is no reason why a nonzero Lorentz term cannot exist above some limiting frequency providing it reduces to zero in the zero frequency limit. In any case, Fejer's argument is solely based on the contribution to the polarization for the electrons, whereas for the case of a nonzero Lorentz term the coupling that exists between electron and ion motion means that both must be considered in the zero frequency limit. This is discussed by Strangeways [1991b] who also investigates the case of a Lorentz term in a plasma containing electrons and ions of finite mass. Further, Fejer does not consider negative l . It is easy to show that, in this case, arguments from causality put no limit at all on the value of negative l , even at zero frequency [Strangeways, 1991b]. Fejer also argued that a Lorentz term would lead to an absurd result for the permittivity of the plasma in the zero frequency limit. This argument also does not apply for negative l but in any case is not rigorous because, as mentioned above, ion motion is not considered.

There are also problems in those proofs that the Lorentz term is zero which consider the perturbations of the orbits of the electrons by the applied field. Darwin [1943] and Theimer and Taylor [1961] showed that these produce an effect equivalent to a force which exactly cancels that due to polari-

zation field. However, these proofs contain an integration of the change in velocity due to the effect of the spatial perturbation on a collision which is taken over all relative orientations of electron velocity to applied electric field direction. The maximum effect is achieved for velocities perpendicular to the direction of the applied field, whereas there is no effect for electron velocity parallel to the applied field. Thus the two forces will only exactly cancel if the weightings are not altered, i.e., if there are an equal number of particles with velocities in each small cone of solid angle. Hence velocity distributions for which this is not true such as a loss cone distribution would be predicted to result in a nonzero Lorentz term which could be either positive or negative. Thus either these proofs are inadequate or a nonzero Lorentz term would exist under such conditions.

A further notable shortcoming of the majority of proofs that the Lorentz polarization term is zero [e.g. Ginzburg, 1970; Kadomtsev, 1958; Darwin, 1943; Theimer and Taylor, 1961] is that they assume the applied electric field is weak. Kadomtsev, for example, assumes a Maxwellian plasma distribution, whereas a strong E field may result in departures from this. If we consider, whistler propagation in ducts, it is evident that the assumption of a weak E field may not be valid. The initiating lightning discharge can result in field strengths in the ionosphere much greater than 0.1 V/m and ducted propagation will serve to prevent appreciable attenuation of the field with distance by limiting the spreading of the wave energy. Calculation shows that the ducted E field can exceed the plasma field and thus an effect of the wave on electron velocity distribution function and electron temperature cannot be ruled out.

Ginzburg [1970] notes the assumption of a weak field outlined above and concludes that the problem of the effective field in a plasma needs further analysis with a more general approach, considering non-linearity and the relation in this problem between the general theory of media with spatial dispersion and the treatment with averaging over a small volume. In the absence of such a treatment, we attempt to consider input to this question from experimental data. Budden [1985] has concluded that the strongest observational evidence that the Lorentz term should not be included is from the theory of whistlers. We thus consider below what effect a nonzero Lorentz term would have on ducted whistler propagation and hence what limits on l can be determined from the experimental data on whistlers. Whereas, it has been previously shown that the lower cutoff of whistlers provides a very low limit ($\sim 10^{-5}$) on positive l [Gershman and Ugarov, 1961], we find that it is the upper cutoff frequency that is the important parameter for negative l . The rather surprising result is obtained that a negative value of l of between about 10^{-4} and 10^{-3} actually provides better agreement with the distribution of whistler cutoffs observed at $L \sim 4$ than the normal assumption that $l=0$. We do not in this paper discuss in detail the physical mechanism that might produce such a nonzero Lorentz term. Although, as mentioned above, a strong E-field is a candidate for this effect,

other explanations cannot be ruled out. In particular, Taylor [1962] showed that an anti-Lorentz term will result for propagation in an isotropic medium with irregularities in electron density smaller than the wavelength. For an anisotropic medium with superposed static magnetic field the situation is more complicated and conflicting results have been obtained [Budden, 1959; Taylor, 1962; Liu, 1967]. In as much as ducts can show considerable fine structure [Mosier, 1976; Koons, 1989], some effect on whistler time delays due to an irregular medium cannot be ruled out. The whole question is really rather involved so that some experimental input, such as provided below, is deemed to be relevant.

3. EFFECT OF A NONZERO LORENTZ TERM ON DUCTED WHISTLER PROPAGATION

If a Lorentz term is included the refractive index n for a neutral plasma with a superposed static magnetic field B , considering only electron motion is given by

$$n^2 = 1 - \frac{X}{(U + lX) - Y_T^2/2(U - mX) \pm [Y_L^2 - 4(U - mX)^2 + Y_T^2]^{1/2}} \quad (4)$$

$X = \omega_p^2/\omega^2$, $Y = \omega_{B,e}/\omega$, $Y_L = Y \cos \psi$ and $Y_T = Y \sin \psi$ where ψ is the angle between the wave normal and the geomagnetic field direction, ω_p is the electron plasma frequency, $\omega_{B,e}$ the electron gyrofrequency, $m = 1 - l$ and $U = 1 - iZ$ where $Z = \nu/\omega$, ν being the collision frequency between electrons and heavy particles.

Substitution in (4) of $l=0$ gives the Appleton formula and $l=1/3$ the Hartree formula, and hence (4) can be correctly termed the Appleton-Hartree formula. Thus the reduced formula obtained by putting $l=0$ in (4) should be called the Appleton formula or the Appleton-Lassen formula. Ratcliffe [1939, 1959] has outlined some of the consequences for the refractive index of taking $l=1/3$. Ratcliffe considered that the theoretical question was so difficult to resolve convincingly that it would be best to investigate experimentally from radio wave experiments whether a Lorentz term existed for ionospheric propagation. Various experimental tests were performed for radio waves reflected from ionospheric layers. Smith [1941], for example, found that for the E layer at 6 MHz, $l = -0.02 \pm 0.05$. A very much smaller limit on the value of $l > 0$ can be inferred for ducted whistler propagation. If we take the usual whistler approximations to the refractive index, then with l nonzero, we obtain

$$n^2 = \frac{f_0^2}{f(f_{B,e} \cos \psi - f) - lf_0^2} \quad (5)$$

Clearly the effect of the Lorentz term on the phase refractive index will be larger at lower frequencies and, for a given $f_{B,e}$ and f_0 , there will be a minimum value of f for the denominator to remain positive. Thus the observed lower cutoff frequency of a whistler can be used to imply a maximum possible value of l . The largest value of the Lorentz term, for a

given l will correspond to the peak of the F2 layer. Let us take $f = 1$ kHz, a reasonable lower cutoff frequency for whistlers, $f_{B,e} = 1.5$ MHz and $f_0 = 5$ MHz. Then, for $\psi = 0^\circ$, we find that $l < 3.10^{-4}$. The above limit is, however, only for l positive since a negative value would not introduce such a cutoff. Thus, the lower cutoff frequency of a whistler places no limit at all on the magnitude of an anti-Lorentz term. However, both positive and negative values will alter the upper cutoff frequency of whistlers for propagation in crests of ionization and can thus provide a negative l limit. Thus we will consider this question below.

Consider a whistler-mode wave propagating in a crest of ionization. For the wave to remain ducted, assuming the duct can be considered to be at least locally uniform along its length, we find applying Snell's law [Smith et al., 1960] that

$$n[N(O), \psi_0] \cos \psi_0 = n[N(P), 0] \quad (6)$$

where $N(O)$ and $N(P)$ are the electron densities at the center and edge of the duct, ψ_0 is the angle of the wave normal to the geomagnetic field direction, and δ is the fractional electron density enhancement, both at the duct center. If we substitute for the refractive index using (5), we obtain

$$\left[\frac{f_0^2(1+\delta)}{f(f_{B,e} \cos \psi_0 - f) - f_0^2(1+\delta)} \right]^{1/2} \cos \psi_0 = \left[\frac{f_0^2}{f(f_{B,e} - f) - f_0^2} \right]^{1/2} \quad (7)$$

Squaring both sides and after some rearrangement:

$$(1+\delta) \cos^2 \psi_0 (ff_{B,e} - f^2 - lf_0^2) = f(f_{B,e} \cos \psi_0 - f) - lf_0^2(1+\delta) \quad (8)$$

The maximum frequency of propagation in a crest of ionization is where the morphology of the refractive index surface changes from concave to convex. Above this frequency a trough of ionization is needed for whistler mode ducting corresponding to $\delta < 0$. Thus the maximum frequency of propagation in a crest f_m is given by $\delta = 0$ in (8), giving

$$\cos^2 \psi_0 [f_m(f_{B,e} - f_m) - lf_0^2] = f_m(f_{B,e} \cos \psi_0 - f_m) - lf_0^2 \quad (9)$$

This maximum frequency will also correspond to $\psi = 0$ so that we can use $\cos \psi_0 = 1 - \psi_0^2/2$ so that

$$(1 - \psi_0^2/2)^2 [f_m(f_{B,e} - f_m) - lf_0^2] = f_m[f_{B,e}(1 - \psi_0^2/2) - f_m] - lf_0^2 \quad (10)$$

neglecting powers of ψ_0 above 2, we obtain

$$f_{B,e} - 2(f_m + lf_0^2/f_m) = 0 \quad (11)$$

If $l = 0$, the normal cutoff for $f_0 \gg f_{B,e}$ is obtained, namely $f_m = f_{B,e}/2$. For $l \neq 0$ the cutoff is altered to

$$f_m = \frac{f_{B,e}}{4} + \frac{f_{B,e}}{4} \left[1 - \frac{16f_0^2}{f_{B,e}^2} \right]^{1/2} \quad (12)$$

and assuming $16lf_0^2 \ll f_{B,e}^2$, we have

$$f_m = 0.5f_{B,e} - \frac{2lf_0^2}{f_{B,e}} \quad (13)$$

Thus the cutoff will be reduced for $l > 0$ and increased for $l < 0$. For whistler propagation, when l is very small, the cutoff will correspond to the minimum electron gyrofrequency along the path which will be at the equatorial plane. Thus the cutoff will be given by:

$$f_m = f_{B,e}(eq)/2 - 2lf_0(eq)^2/f_{B,e}(eq) \quad (14)$$

Strangeways [1991a] considers in detail the upper cutoff frequency of whistlers as measured experimentally and as would be predicted on the basis of ducted propagation in crests of ionization. The latter is determined by ray tracing in a realistic magnetospheric model including field-aligned ducts of Gaussian cross section. The ray tracing model represented winter night conditions and contained electrons and hydrogen, helium and oxygen ions in diffusive equilibrium [Denby et al., 1980]. The enhancement Δ of a superposed duct at any position across its width was given by:

$$\Delta = \delta \exp(-x^2/2\sigma^2) \quad (15)$$

where x is the separation, in the equatorial plane, between field lines corresponding to the ray position and the duct center, δ is as defined above, and σ is the distance either side of the duct center where the density enhancement has fallen to $\delta/e^{1/2}$. The solid line in Figure 2 reproduces the cutoff as a function of duct enhancement for a duct of Gaussian cross-section with $\sigma = 100$ km, corresponding to typical duct widths observed by satellite-borne receivers (Smith and Angerami, 1968; Angerami, 1970; Koons, 1989). Rays were started with field-aligned wave normals at 1000 km altitude and were considered to be ducted if still in the duct when they reached the same altitude in the conjugate hemisphere. The figure shows that the enhancement required to duct rays at a frequency corresponding to $0.5f_{B,e}(eq)$ is significantly larger than at lower frequencies and a few times larger than found previously by Angerami [1970] or Laird [1981] using more approximate treatments [c.f. Strangeways, 1991a]. A mode theory treatment [Karpman and Kaufman, 1984] for typical whistler parameters reveals mode cutoffs for frequencies in the range 0.44 and 0.49 in the equatorial region, with 10 modes at $f = 0.4f_{B,e}$ reducing at higher frequency to only one mode which cuts off at $f = 0.488f_{B,e}$. Thus a mode theory treatment shows that, for density enhancement ducts, all cutoffs will be less than $0.5f_{B,e}$. The ray tracing calculations (Figure 2) show that for smaller enhancements, lower values of the cutoff result. This is caused by the transverse gradient of electron gyrofrequency across the duct and the effect of the curvature of the geomagnetic field lines as discussed in detail by Strangeways [1991a]. No ray paths were found to have a cutoff above $0.5f_{B,e}(eq)$. Although it is possible to trace such paths using rather special initial conditions [Bernhardt, 1979] or by rely-

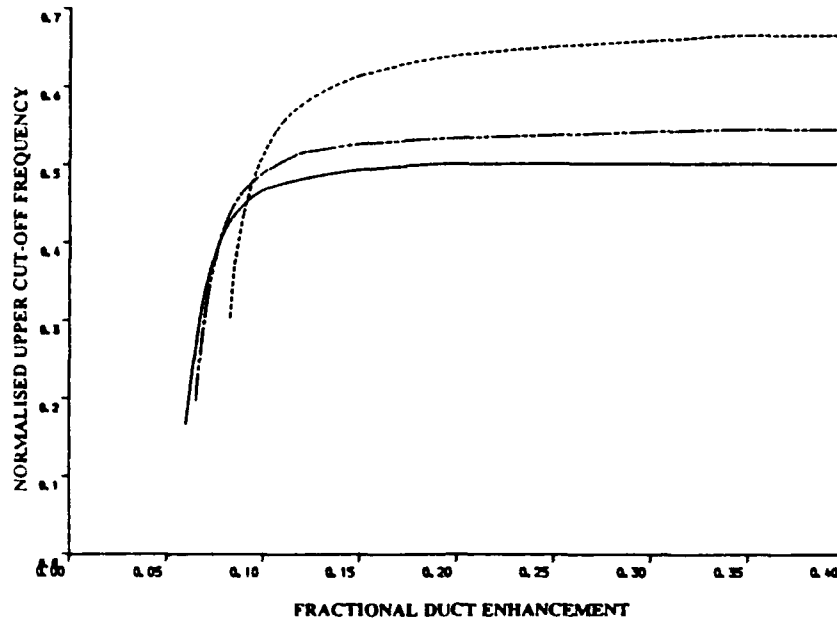


Fig. 2. The variation of whistler upper cutoff frequency normalized to the actual equatorial electron gyrofrequency for a $\sigma = 100$ -km duct in the WN model as a function of duct enhancement for $l=0$ (solid line), $l = -10^{-4}$ (dashed-dotted line), $l = -4 \times 10^{-4}$ (dashed line).

ing on an outer or inner duct to reflect the ray back into the duct [Strangways, 1982], such paths do not really represent a higher whistler cutoff as the amount of energy propagating by them would be much reduced from that for frequencies below $0.5 f_{B,e}(eq)$. This can be seen by comparing the number of inferred ducted ray paths that exist for $f > 0.5 f_{B,e}(eq)$ compared with $f < 0.5 f_{B,e}(eq)$. Such a comparison can be made from Figures 6 and 7 of Bernhardt [1979]. The author has found the same sort of result for the two-duct trapping mechanism described by Strangways [1982]; as the frequency is increased above $0.5 f_{B,e}(eq)$, the range of initial latitude for which inter-hemispheric ducted paths exist becomes smaller and smaller. These results therefore correspond to a more gradual cutoff still centered on $0.5 f_{B,e}(eq)$. Experimental results, however, show an abrupt cutoff for many whistlers at frequencies a little above $0.5 f_{B,e}(eq)$. What is required is an explanation of how the normal cutoff can be increased rather than a way by which the cutoff can be made more gradual so that a relatively few ray paths can be traced for $f > 0.5 f_{B,e}(eq)$. Carpenter [1968] measured the cutoff of 541 whistler traces which were mostly recorded at Eights, Antarctica from June to August, 1963 and found $R_{co} = f_n/f_{B,e}(eq) = 0.51 \pm 0.03$. Thus more than 50% of the whistlers had a cutoff higher than permitted by the normal theory of whistler-mode ducting in crests and as found by the ray tracing calculations presented above for $l=0$. Carpenter's results are not atypical. Liemohn and Scarf [1964], for example, obtained a similar distribution for 57

nose whistlers recorded at College in Alaska, Seattle and Byrd station, Antarctica (assuming an R^{-3} electron density distribution). It is clear from (14) above that such higher cutoff frequencies could be explained by a negative value of l , i.e. an anti-Lorentz term which is equivalent to a restorative force acting on the electrons. It might be argued that these higher cutoffs could easily be explained in terms of some more conventional effect. It is possible, for instance, that they might result from some effect of the wave particle interaction causing their amplification. However, calculations of the effect of cyclotron resonance show a negligible effect on the real part of the refractive index [Matsumoto and Kimura, 1971] so that there is not at present any credible theory for such an effect. A number of authors have investigated the effect of a warm or hot plasma on whistler-mode propagation; (Guthart, 1965a; Hashimoto et al., 1977; Moreira, 1982; Sazhin et al. 1990). Sazhin [1987] has also considered propagation in a weakly relativistic plasma. It can be concluded from these treatments that for ducted propagation (ψ small) inside the plasmasphere and for realistic values of equatorial electron temperature and electron density fall-off with radial distance, there will be no significant increase in the ratio of whistler upper cutoff frequency to estimated equatorial electron gyrofrequency due to temperature or relativistic effects.

It could also be argued that the cutoffs above $0.5 f_{B,e}(eq)$ are not real but rather result from using an incorrect relationship between f_n and $f_{B,e}$ when determining the equatorial

electron gyrofrequency from the nose frequency of the whistlers. The relation $f_n = 0.38f_{B,e}(eq)$ used by Carpenter [1968] was based on Angerami's [1966] work on the diffusive equilibrium model; it could perhaps be argued that the constant 0.38 was not correct. Park [1972] calculated this constant for field-aligned whistler propagation for four different diffusive equilibrium models (corresponding to paths inside the plasmasphere). At $L=4$ the constant was found to vary between 0.384 and 0.408, showing that it was not particularly sensitive to the plasma composition and temperature which varied amongst the models. For a more rapid electron density decrease along the field line, corresponding to a collisionless plasma, the constant was 0.435. For the experimentally observed cutoffs to have been overestimated requires a lower value of this constant than 0.38. This therefore seems very unlikely.

Various authors have considered Landau damping as a mechanism to explain whistler cutoffs [Scarf, 1962; Liemohn and Scarf, 1964; Guthart, 1965b; Liemohn, 1967], but clearly no damping mechanism can increase a cutoff limit set by propagational effects (i.e. duct loss at $0.5f_{B,e}(eq)$) so that this mechanism cannot increase the cutoff frequency for ducting in crests of ionization. Various other possible explanations are discussed by Strangways [1991a] and shown to be unsatisfactory.

4. RAY TRACING CALCULATIONS WITH AN ANTI-LORENTZ TERM

In this section we investigate whistler-mode ducting in Gaussian cross-section ducts in the winter night (WN) model of the magnetospheric plasma with an anti-Lorentz term added to see whether this can adequately explain the observed upper cutoff frequencies of whistlers above $0.5f_{B,e}(eq)$. We will only consider the cutoffs of normal whistlers. We consider that super-whistlers [Bernhardt, 1979] which occur only very rarely and have cutoffs above 0.6 times $f_{B,e}(eq)$ (determined by curve fitting to their spectra) require a different mechanism for their explanation. A Lorentz term with positive l would result in cutoffs below $0.5f_{B,e}(eq)$. As there are many other possible explanations for a decrease in cutoff and since, in general, whistler cutoffs are observed to be larger than expected, we do not consider the case of the effect of positive l on whistler upper cutoffs. The refractive index used in the ray tracing program is that given in (4), taking $\nu=0$. Derivatives of n with respect to the spatial coordinates (r, θ) were determined for the Haselgrove [1954] ray tracing equations. The cutoff as a function of duct enhancement can then be determined in the same manner as Strangways [1991a] and plotted in the same form. The result of this for the $\sigma = 100$ km duct centered on $L=4$ is also shown in Figure 2 for values of l of -10^{-4} and -5×10^{-4} . It is clear that cutoffs above $0.5f_{B,e}(eq)$ are possible for both values of l . For a 15% enhancement duct the cutoff is 7.2 kHz, corresponding to $0.527f_{B,e}(eq)$ for $l = -10^{-4}$ and 8.4 kHz, corresponding to $0.615f_{B,e}(eq)$ for $l = -5 \times 10^{-4}$ whereas it was 6.75 kHz

($0.498f_{B,e}(eq)$) for $l=0$. These values are normalized to the actual equatorial gyrofrequency at $L=4$ in the WN model. However, for a proper comparison with the cutoffs for ground-observed whistlers, mentioned above, the cutoff should be normalized to the equatorial electron gyrofrequency that would be observed by curve-fitting to the whistlers. This is achieved by first determining the $f-t$ relationship for a whistler for a given value of l by ray-tracing. The time delays are obtained in 500 Hz steps from 2 kHz to the cutoff frequency with an additional time delay taken at the actual cutoff. Curve-fitting is then performed to these theoretical whistler time delays over their entire frequency range using the curve-fitting program described by Tarcasai [1975]. This method has been proved to be reliable for the analysis of both real and theoretically derived nose and non-nose whistlers and is based on Bernard's [1973] formula for the dispersion $D(f) \approx t \cdot f^{1/2}$ which approximates the travel time integral with an analytic expression. The method uses a least squares estimation of the zero frequency dispersion D_0 , the equatorial electron gyrofrequency $f_{B,e}(eq)$ in Bernard's approximation, and a third parameter T , locating the time of the causative atmospheric. The program minimizes the sum of the squares of the differences between the observed and calculated propagation times. After obtaining these three parameters the L value of propagation, the electron density at 1000 km and at the equator, and the tube electron content are also calculated using the quasi-constants given by Park [1972]. In this way, the equatorial gyrofrequency that would be inferred from the whistler $f-t$ relationship for a given l can be determined. Figure 3 shows the normalized upper cutoff frequency as a function of $-l$. The continuous line shows the result when the upper cutoff frequency is normalized to the actual value of $f_{B,e}(eq)$ whereas the dashed line shows the result of normalizing to the equatorial electron gyrofrequency that would be obtained from curve fitting to the whistler $f-t$ values. The latter corresponds to what would be measured for ground-observed whistlers. The difference arises because the anti-Lorentz term increases the nose frequency a little as well as the cutoff. The maximum cutoff frequency for any value of l is about $0.55f_{B,e}(eq)$. The lowest cutoff is about 0.49 but smaller cutoffs would result for duct enhancements less than 15% or for ducts of 15% enhancement but with greater width. Thus the range of cutoffs that would be expected to be observed for a reasonable range of duct enhancements and widths is in good agreement with experimental results such as those presented by Carpenter [1968], whereas the assumption that $l=0$, which is generally made, results in a maximum cutoff value of $0.5f_{B,e}(eq)$ and thus a distribution centered on too low a cutoff frequency. Thus the inclusion of an anti-Lorentz term ($l < 0$) appears to provide better agreement with the experimental data than taking $l=0$. However, before we can conclude that the upper cutoff frequencies of whistlers imply an anti-Lorentz term for ducted whistler propagation, it is necessary to investigate whether the presence of an anti-Lorentz term

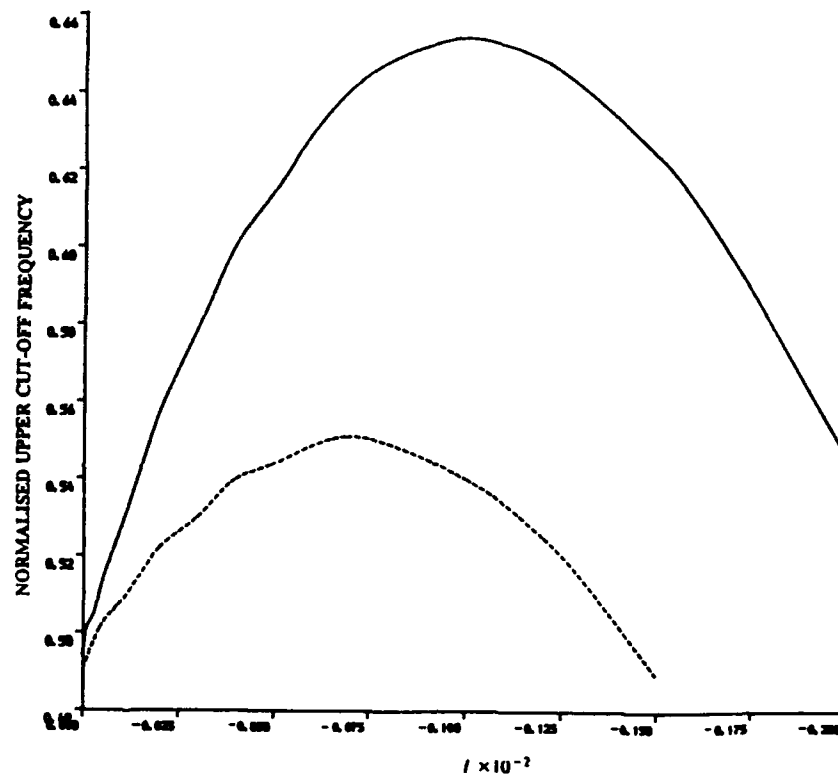


Fig.3. The variation of whistler cut-off frequency for a 15% enhancement $\sigma = 100$ km duct in the WN model; normalized to the equatorial electron gyrofrequency (solid line); normalized to the equatorial electron gyrofrequency that would be determined by curve-fitting to the whistler f - t values (dashed line).

would severely distort the f - t relationship for whistlers in addition to increasing the cutoff. We first note that the ray tracing calculations showed that there was a maximum negative value of l of about -2×10^{-3} for ducting to be possible in the WN model. This limit will be very significant if $|l|$ were to increase with wave amplitude as it would provide a saturation mechanism for amplification of a ducted wave by cyclotron resonance. Numerical modeling of wave-particle interactions at whistler frequencies show that such a loss mechanism is necessary for stability and to limit wave growth. For values of l close to this limit there is found to be both an upper and lower cutoff frequency.

Figure 4 shows the time delay versus frequency from 2 kHz to cutoff for five different values of l . These are $l=0$, -10^{-4} , -2×10^{-4} , -5×10^{-4} , and -10^{-3} . Ray paths were not traced at lower frequencies than 2 kHz because the effect of ions is not included in the formula for the phase refractive index (equation (5)). The whistler dispersion appears very similar to that of observed whistlers for all cases, except for $l = -10^{-3}$. It might at first sight seem surprising that the spectra are so little altered from the $l=0$ case at low frequencies. Considering the phase refractive index approximation given by (5), it might be expected that the greatest

difference for nonzero l would occur at the lowest frequencies where the Lorentz term is most comparable with the other term in the denominator. The fact that this is not the case is explained by considering the expression for the group refractive index μ' with $l \neq 0$. Using (5) for n and taking $\psi=0$, we obtain

$$\mu' = \frac{f_0(f f_{B,e} - 2lf_0^2)}{2[f(f_{B,e} - f) - lf_0^2]^{3/2}} \quad (16)$$

A term involving l also appears in the numerator of the expression for μ' ; this can compensate for the increase in the denominator when $l < 0$ for some frequencies below the nose.

A more stringent test of the dispersion is to examine the residuals after curve fitting to the f - t values using Tarcsai's program. The residual at any frequency is the time delays calculated by ray tracing minus the calculated value at the same frequency for the least squares fit to Bernard's formula over the whole whistler frequency range [cf. Tarcsai, 1975; Tarcsai et al., 1989]. For negative values of l up to -4×10^{-4} , the residuals at all frequencies were always less than 1.5 ms; such values are fairly insignificant as they are less than typical scaling errors for whistlers ducted at $L=4$. For more negative values of l the largest residual at any frequency was

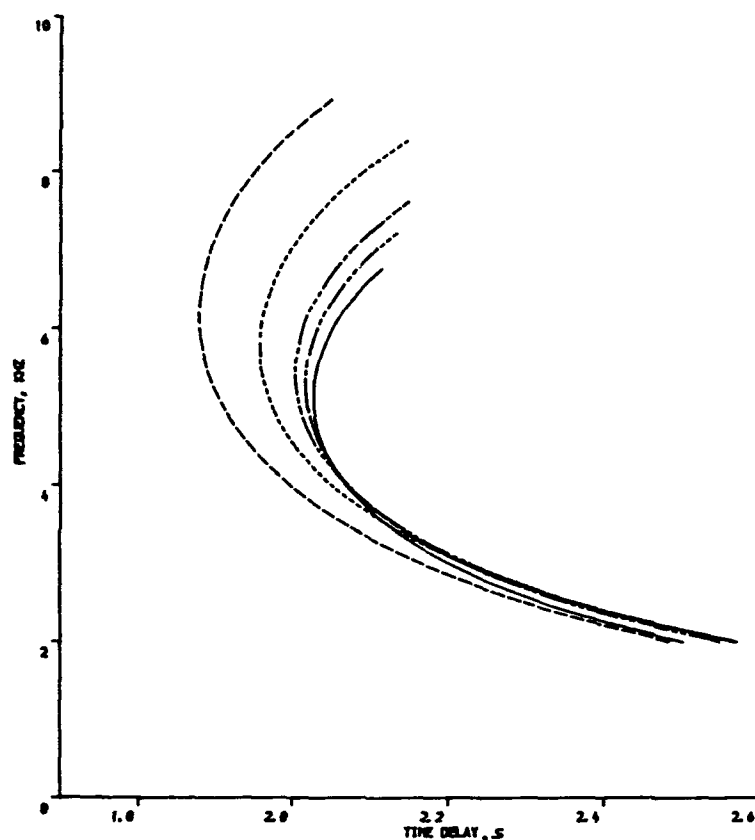


Fig.4 Ray traced whistlers for $l=0$ (solid line), $l = -10^{-4}$ (dashed-dot-dot line), $l = -2 \times 10^{-4}$ (dashed-dot-dot-dot line), $l = -5 \times 10^{-4}$ (short dashed line), $l = -10^{-3}$ (long dashed line).

10.5 ms; this was at 2 kHz for $l = -1.2 \times 10^{-3}$. At 2 kHz at $L=4$ the presence of ions introduces an additional time delay of about 20 ms. Also, propagation in the Earth-ionosphere waveguide will introduce an extra time delay for $f \approx 2$ kHz, near the Earth-ionosphere waveguide cutoff. These effects might compensate for the reduced delay due to negative l for real whistlers. The residuals calculated would all be smaller if they were not obtained by curve fitting to so many points (about 15) or over such a wide frequency range (about 6 kHz). Nevertheless the residuals are fairly small (less than about 3 ms) for small l ($< 8 \times 10^{-4}$). The residuals would also be significantly smaller for any given value of l for a plasmasphere model with a faster rate of falloff of electron density along the field line as would occur if a realistic temperature gradient were incorporated [Strangeways, 1986] or if a collisionless rather than diffusive equilibrium model were considered. For the WNT1 model [Strangeways, 1986] in which the electron temperature is taken to be proportional to the radial distance squared, residuals were all found to be less than 1.5 ms for $l = -1.2 \times 10^{-3}$ and ducting was found to still occur for $l = -10^{-2}$. We conclude that the possibility of an anti-Lorentz term (or some other term giving a similar

effect) for ducted whistler propagation must be seriously considered.

5. CONCLUSIONS

It has been shown that a nonzero Lorentz term cannot be ruled out on the basis of theoretical arguments except for positive l at zero frequency and that the cutoff frequencies of nose whistlers observed at $L \sim 4$ are in better agreement with propagation in a medium for which there is a small anti-Lorentz term ($l \sim -10^{-4}$ to -10^{-3}) than one for which the Lorentz term is assumed to be zero. Whistler spectra are not significantly distorted except for larger values of $-l$: the value of $-l$ for which noticeable distortion occurs depends on the equatorial electron density. The distribution of upper cutoff frequencies of whistlers received on the ground at $L \sim 4$ can be explained by either a variation in $-l$ or in duct enhancement or both.

REFERENCES

- Angerami, J. J., A whistler study of the distribution of thermal electrons in the magnetosphere, *Tech. Rept. 3412-7*,

- Radioscience lab., Stanford University, Stanford, Calif., 1966.
- Angerami, J. J. Whistler duct properties deduced from VLF observations made with the OGO 3 satellite near the magnetic equator, *J. geophys. Res.*, **75**, 6115 - 6135, 1970.
- Appleton, E. V., Wireless studies of the ionosphere, *J. Inst. Electr. Eng.*, **71**, 642-650, 1932.
- Bernard, L. C., A new nose extension method for whistlers, *J. atmos. terr. Phys.*, **35**, 871-880, 1973.
- Bernhardt, P. A., Theory and analysis of the "Super-whistler", *J. geophys. Res.*, **84**, 5131-5142, 1979.
- Budden, K. G., Effect of small irregularities on the constitutive relations of the ionosphere, *J. natl. res. bur. Stand., Sect. D*, **63**, 135-149, 1959.
- Budden, K. G. The Propagation of Radio Waves, pp.40-42, Cambridge University Press, New York, 1985.
- Carpenter, D. L., Ducted whistler-mode propagation in the magnetosphere; a half gyrofrequency upper intensity cut-off and some associated wave growth phenomena, *J. geophys. Res.*, **73**, 2919 - 2928, 1968.
- Chelkowski, A., Dielectric physics, Appendix A, Elsevier, New York, 1980.
- Darwin, C. G., The refractive index of an ionised medium, *Proc. R. Soc. London Ser. A*, **182**, 152 - 166, 1943.
- Denby M., K. Bullough, P. D. Alexander, and M.J. Rycroft, Observational and theoretical studies of a cross meridian refraction of VLF waves in the ionosphere and magnetosphere, *J. atmos. terr. Phys.*, **42**, 51-59, 1980.
- Fejer J. A., Causality and the Lorentz polarisation term, *J. atmos. terr. Phys.*, **47**, 513-516, 1985.
- Gershman, B.N. and V.A. Ugarov, Propagation and generation of low-frequency electromagnetic waves in the upper atmosphere, *Sov. Phys. Usp.*, Engl. Transl., **3(5)**, 743 - 763, 1961.
- Ginzburg V.L., The Propagation of Electromagnetic Waves in Plasmas, 2nd ed., pp.22-25, Pergamon, New York, 1970.
- Guthart H., Nose whistler dispersion as a measure of magnetospheric electron temperature, *J. res. natl. bur. Stand., Sect. D*, **69**, 1417-1424, 1965a.
- Guthart H., An anisotropic electron velocity distribution for the cyclotron absorption of whistlers and VLF emissions, *J. res. natl. bur. Stand., Sect. D*, **69**, 1403-1415, 1965b.
- Hartree D. R., The propagation of electromagnetic waves in a stratified medium, *Proc. Cambridge Philos. Soc.*, **25**, 97-120, 1929.
- Hashimoto K., I. Kimura, and H. Kumagai, Estimation of electron temperature by VLF waves propagating in directions near the resonance cone, *Planet. Space Sci.*, **25**, 871-878, 1977.
- Haselgrove, J. Ray theory and a new method for ray tracing, report of the conference on the physics of the ionosphere, pp.335 - 364, London Phys. Soc., 1954.
- Kadomtsev, B. B., On the effective field in a plasma, *Sov. Phys. JETP*, Engl. Trans., **6**, 117, 1958.
- Karpman, V. I. and R. N. Kaufman, Whistler wave propagation in magnetospheric ducts (in the equatorial region), *Planet. Space Sci.*, **32**, 1505 - 1511, 1984.
- Koons, H. B. Observations of large amplitude whistler mode wave ducts in the outer plasmasphere, *J. geophys. Res.*, **94**, 15,393 - 15,397, 1989.
- Lassen H. Über den Einfluss des Erdmagnetfeldes auf die Fortpflanzung der elektrischen Wellen der drahtlosen Telegraphie in der Atmosphäre, *Electr. Nachr.-Technik*, **4**, 324 - 334, 1927.
- Laird, M. J. Reflection of electromagnetic waves by density gradients in a magnetised plasma, *J. atmos. terr. Phys.*, **43**, 81-86, 1981.
- Liemohn, H. B. Cyclotron resonance amplification of VLF and ULF whistlers, *J. geophys. Res.*, **72**, 39 - 55, 1967.
- Liemohn, H. B. and F. L. Scarf Whistler determination of electron energy and density distributions in the magnetosphere, *J. geophys. Res.*, **69**, 883-904, 1964.
- Liu, C. H. Effective dielectric tensor and propagation constant of plane waves in a random anisotropic medium, *J. math. Phys.*, **8**, 2236 - 2242, 1967.
- Lorentz, H. A., The Theory of Electrons, pp.138-139 and 308, B.G. Teubner, Leipzig, Germany, 1909.
- Matsumoto, H., and I. Kimura, Linear and nonlinear cyclotron instability and VLF emissions in the magnetosphere, *Planet. Space Sci.*, **19**, 567-608, 1971.
- Moreira, A., On the influence of a plasma hot component on whistler propagation beyond the plasmopause, *J. atmos. terr. Phys.*, **44**, 869 - 874, 1982.
- Mosier, S. R., Observations of magnetospheric ionization enhancements using upper hybrid resonance noise band data from RAE 1 satellite, *J. geophys. Res.*, **81**, 253-256, 1976.
- Park, C. G., Methods of determining electron concentration in the magnetosphere from nose whistlers, technical report, Stanford Radiosci. Lab., Stanford Univ., Stanford, Calif., 1972.
- Ratcliffe, J. A. 1939 The effect of the Lorentz polarisation term in ionospheric calculations, *Proc. Phys. Soc.*, **51**, 747-756, 1939.
- Ratcliffe, J. A. The Magneto-ionic Theory and its Applications to the Ionosphere, p.154 - 164, Cambridge University Press, New York, 1959.
- Sazhin, S.S., An approximate treatment of electromagnetic wave propagation in a weakly relativistic plasma, *J. Plasma Phys.*, **37(2)**, 209-230, 1987.
- Sazhin S. S., A. J. Smith, and E. M. Sazhina, Can magnetospheric electron temperature be inferred from whistler dispersion measurements?, *Ann. Geophys.*, **8(4)**, 273 - 285, 1990.
- Scarf, F. L., Landau damping and attenuation of whistlers, *Phys. Fluids*, **5**, 6-13, 1962.
- Smith, N., Oblique incidence radio transmission and the Lorentz polarisation term, *J. Res. Natl. Bur. Stand., Sect. D*, **26**, 105 - 116, 1941.

- Smith, R. L., and J. J. Angerami, Magnetospheric properties deduced from OGO 1 observations of ducted and non-ducted whistlers, *J. geophys. Res.*, 73, 1-20, 1968.
- Smith, R. L., R. A. Helliwell and I. W. Yabroff, A theory of trapping of whistlers in field-aligned columns of enhanced ionization, *J. geophys. Res.*, 65, 815 - 823, 1960.
- Strangeways, H. J. The effect of multi-duct structure on whistler-mode propagation, *J. atmos. terr. Phys.*, 44, 901-912, 1982.
- Strangeways, H. J., A model for the electron temperature variation along geomagnetic field lines and its effect on electron density profiles and VLF ray paths, *J. atmos. terr. Phys.*, 48, 671 - 683, 1986.
- Strangeways, H. J., The upper cut-off frequency of nose whistlers and implications for duct structure, *J. atmos. terr. Phys.*, 53, 151 - 169, 1991a.
- Strangeways, H. J., The Lorentz polarisation term for ionospheric and magnetospheric e-m propagation, in press, *J. atmos. terr. Phys.*, 1991b.
- Tarcsai, Gy., Routine whistler analysis by means of accurate curve fitting, *J. atmos. terr. Phys.*, 27, 1447 - 1457, 1975.
- Tarcsai, Gy., H. J. Strangeways and M.J. Rycroft, Error sources and travel time residuals in plasmaspheric whistler interpretation, *J. atmos. terr. Phys.*, 51, 249 - 258, 1989.
- Taylor, L. S., The index of refraction of an irregular ionized gas, *J. geophys. Res.*, 67, 3843- 3849, 1962.
- Taylor, L. S., Radio waves in turbulent magnetoplasmas, *Phys. Fluids.*, 8, 492- 499, 1965.
- Theimer O., and L. S. Taylor, On the refractive index of the ionosphere, *J. geophys. Res.*, 66, 3157- 3162, 1961.

H. J. Strangeways, Department of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, W. Yorks, U.K.

On the location of causative atmospherics of very low latitude whistlers and their magnetospheric propagation mechanism

M. Hayakawa¹

Solar-Terrestrial Environment Laboratory, Nagoya University, Toyokawa, Aichi, Japan

S. Shimakura and M. Moriizumi

Department of Electrical and Electronic Engineering, Chiba University, Chiba, Japan

K. Ohta

Department of Electronic Engineering, Chubu University, Kasugai, Aichi, Japan

(Received January 7, 1991; revised May 1, 1991; accepted June 18, 1991.)

The purpose of this paper is to locate the causative atmospherics of very low latitude whistlers whose ionospheric exit regions have previously been determined at geomagnetic latitudes, 10° – 13° , in order to aid our understanding of the magnetospheric propagation mechanism of very low latitude whistlers. It is found that the causative atmospherics of very low latitude whistlers are very widely distributed in a range from 500 to ~2500 km from the conjugate point of whistler ionospheric exit region previously determined. This fact lends further support to our previous implication that magnetospheric preferred paths supporting field-aligned propagation are extremely rare in occurrence and that there has existed a preferred and stable magnetospheric path in a very restricted location on the day of high whistler activity.

1. INTRODUCTION

The simultaneous determination of the ionospheric exit region of whistlers and of their causative atmospherics is known to be important in the extensive study of the propagation mechanism of whistlers along the total path from the source to the receiver [Hayakawa and Tanaka, 1978]. The direction finding to locate the ionospheric exit regions of whistlers has been currently used in investigating the propagation characteristics of whistlers, whistler ducts, magnetospheric plasma dynamics, etc. [Bullough and Sagredo, 1973; Leavitt, 1975; Okada et al., 1977, 1981; Ohta et al., 1984; Shimakura and Hayakawa, 1987], but the simultaneous location of the ionospheric exit regions of whistlers and their causative atmospherics has so far been lacking [Shimakura et al., 1991], which would be of great importance in the study of very low latitude whistlers.

¹ Now at University of Electro-Communications, Chofu, Tokyo, Japan.

Copyright 1992 by the American Geophysical Union.

Paper number 91RS02113.
0048-6604/92/91RS-02113\$08.00

Hayakawa and Tanaka [1978] have made a comprehensive review of the propagation of lower-latitude whistlers, and they have concluded that the whistlers observed at geomagnetic latitudes less than 20° are very poorly understood. A pioneering contribution in this field has been made by Ondoh et al. [1979], who have indicated remarkable differences of very low latitude whistlers from mid-latitude ones including the nearly overhead ionospheric exit points, the simultaneous observation at Okinawa (geomagnetic latitude 15.3°) and Ishigaki (12.9°), whistler echoes, etc. Further, they have suggested ducted propagation, but later works such as Andrews [1979], Liang et al. [1985], and Thomson [1987] have arrived at the conclusion contradictory to Ondoh et al.'s (i.e., nonducted propagation). The latest observation based on the spaced direction finding by Hayakawa et al. [1990] has indicated that many observational facts are satisfactorily interpreted in terms of the field-aligned propagation as Ondoh et al. suggested, though the field-aligned irregularities are considered to be of filamentary structure, seeming to be different from the ordinary ducts by Ondoh et al. In this paper we deal with the causative atmospherics of very low latitude whis-

tlers as well as those whistlers observed at very low latitudes (path geomagnetic latitude $\approx 10^\circ$). The ionospheric exit regions of very low latitude whistlers have already been determined by means of the spaced direction finding measurements made in South China [Hayakawa *et al.*, 1990], and so we pay additional particular attention to the location of their causative atmospherics in this paper. On the basis of our previous whistler direction finding results [Hayakawa *et al.*, 1990] and the present analysis of their causative atmospherics, we will be able to determine whether the present analysis of causative atmospherics will provide further evidence for the presence of preferred magnetospheric paths and field-aligned propagation of very low latitude whistlers in the inner plasmasphere as suggested in the paper of Hayakawa *et al.* [1990].

2. LOCATIONS OF CAUSATIVE ATMOSPHERICS OF VERY LOW LATITUDE WHISTLERS

The measurements of VLF wideband (2–6 kHz) signals were made simultaneously during January 5–11, 1988, at three stations in South China, (1) Zhanjiang (abbreviated at ZJ; geomagnetic latitude, 10.0°), (2) Guilin (GL; 14.1°), and (3) Wuchang (WC; 19.4°), and three field components (two horizontal magnetic and one vertical electric) are recorded [Hayakawa *et al.*, 1990]. The same digital recorders were adopted at ZJ and WC, while an analogue recording was used at GL, and hence we utilize the data from the two stations of ZJ and WC in this paper.

The whistler events on several days have been analyzed in detail in the paper of Hayakawa *et al.* [1990], but in this paper we deal only with two of them, January 5 and 6. On January 5 there was an enhanced occurrence peak of whistlers (dispersion (D) = $10.5 \text{ s}^{1/2}$) just before LT = 3 hours and the ionospheric exit regions of many whistlers during this occurrence peak are found to be just overhead of the station of ZJ (geomagnetic latitude $\sim 10^\circ$) on the basis of direction finding [Hayakawa *et al.*, 1990]. The causative atmospherics of those whistlers have been located by means of the method proposed by Shimakura *et al.* [1990] and Shimakura and Moriizumi [1990] in which the dispersion effect of atmospherics is used to deduce the propagation distance and ionospheric height with surprisingly

high accuracy. The locations of causative atmospherics of such very low latitude whistlers during the period of LT = 2 hours, 40 min to 42 min have been determined in Figure 1 as black dots, and the conjugate point of the ionospheric exit region of whistlers is also indicated by an open circle. This conjugate point was determined by using the real magnetic field model of International Geomagnetic Reference Field (IGRF) (1985). For the sake of comparison as done by Yoshino [1976], the distribution of thunderclouds observed at the time (LT = 2 hours) closest to the whistler occurrence peak time, is given in Figure 2. This is obtained by picking up the thunderclouds which appear to be whitish bright spots or smaller circular clouds in the infrared cloud picture taken by the Japanese meteorological satellite, because Pierce [1969] has indicated that thunderstorms typically occur within complexes of diameter 30–40 km, and in any complex there are one or two electrically active cells of diameter about 6 km. The figure shows that there are so many thunderclouds, and so it is rather difficult to make one-to-one correlations between the atmospherics in Figure 1 and the corresponding thunderclouds indicated by dark shading in Figure 2. However, there is a tendency that the atmospherics located in Figure 1 are distributed within the regions of many thunderclouds. A small difference in the location of atmospherics and thunderclouds may be attributed to a time difference of about 40 min in the observations of VLF waves and clouds. As seen from Figure 1, the causative atmospherics of whistlers which have traversed the ionosphere near ZJ were widely distributed, and their maximum distance from the conjugate point of the whistler exit regions is about 1500 km.

We have performed the same analyses for the second event on January 6. During the occurrence peak of whistlers just after LT = 1 hour, the ionospheric exit region of whistlers ($D = 13.0 \text{ s}^{1/2}$) was located to about 600 km east of ZJ and WC as indicated by an open circle in Figure 3 [Hayakawa *et al.*, 1990]. The locations of the causative atmospherics of such very low latitude whistlers (geomagnetic latitude $\sim 13^\circ$) during the whistler occurrence peak (LT = 1 hour, 10 min – 1 hour, 3 min) are plotted in Figure 3 again by solid dots, which are confirmed to be consistent with the cloud distribution by the meteorological satellite (although not shown in this paper). The conjugate point of whistler exit regions is indicated as well in the figure as

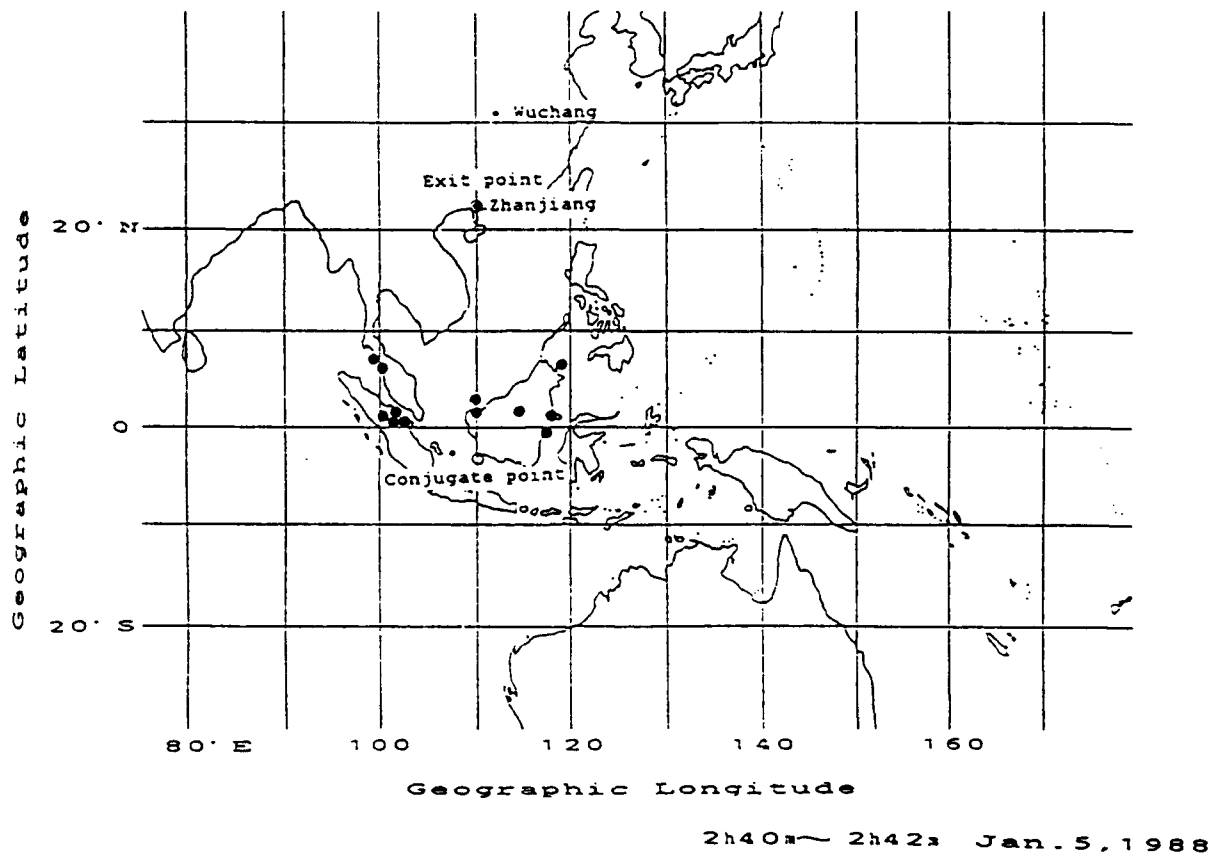


Fig. 1. The source locations of causative atmospherics of very low latitude whistlers during LT = 2 hours. 40 min ~ 2 hours, 42 min on January 5, 1988. The causative atmospherics are indicated by dots, and also an open circle indicates the conjugate of the whistler exit regions.

an open circle. The distance of the causative atmospherics from the conjugate point of the ionospheric exit region of resultant whistlers is found to be in a wide range from about 500 to ~2500 km.

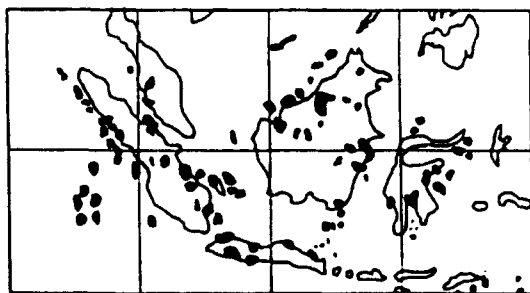


Fig. 2. The distribution of thunderclouds observed by the Japanese meteorological satellite at LT = 2 hours.

3. DISCUSSION ON THE MAGNETOSPHERIC PROPAGATION OF VERY LOW LATITUDE WHISTLERS

The data on both the ionospheric exit regions of whistlers and the location of their causative atmospherics are used here to discuss the propagation mechanism of very low latitude whistlers along the total path from the source to the receiver. As seen from Figures 1 and 3, we have already found that the ionospheric exit region of whistlers on January 5 is located immediately above the station of ZJ, while that on January 6 is located about 600 km east of the two stations of GL and ZJ. Then, Hayakawa *et al.* [1990] have concluded that many characteristics of very low latitude whistlers are satisfactorily interpreted in terms of field-aligned propagation and such field-aligned preferred paths are restricted to a

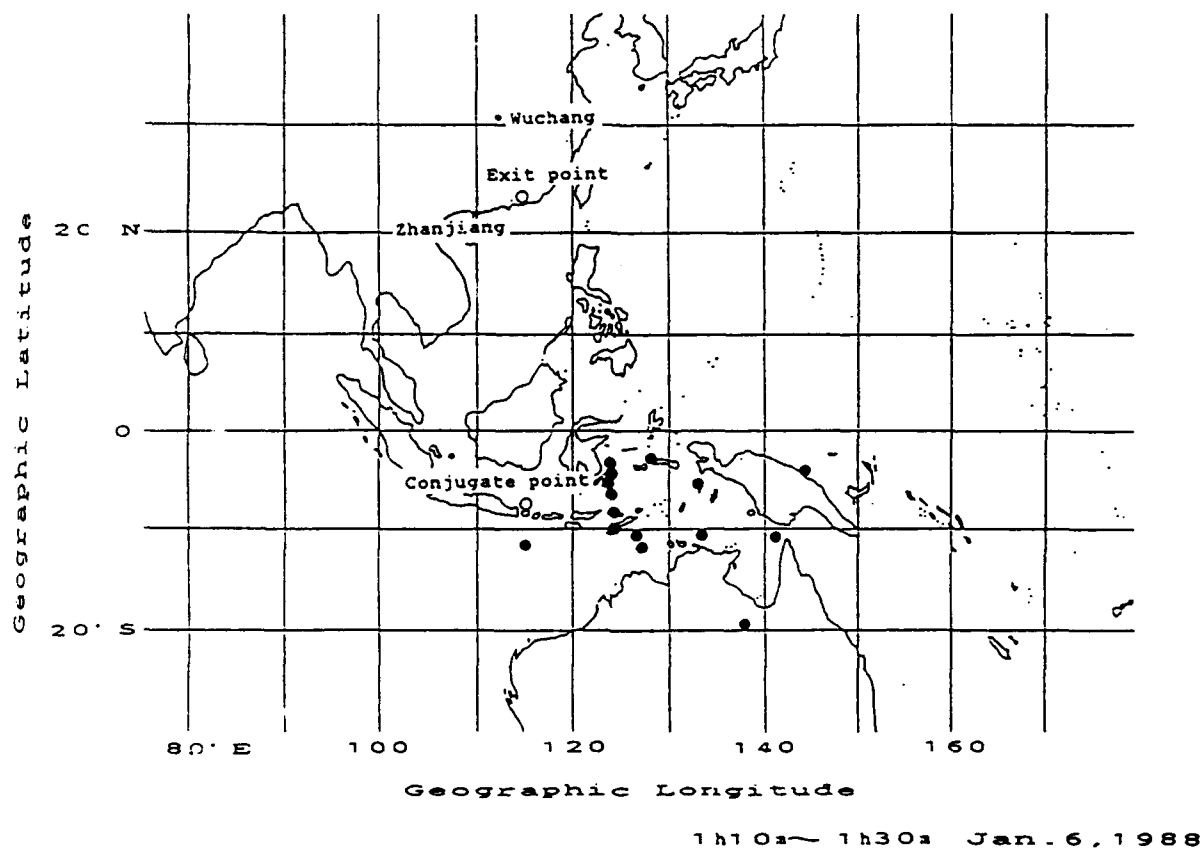


Fig. 3. The same as Figure 1 except during LT = 1 hour, 10 min - 1 hour, 30 min on January 6.

range in geomagnetic latitude (10° – 13°). Hence a preferred field-aligned path in the magnetosphere which existed on both days, is found to accept atmospherics generated over a wide area, leading to the generation of whistlers. Atmospherics on January 5 within a distance of 1500 km from the conjugate point of the whistler exit region are trapped by field-aligned path whose exit region is overhead of ZJ, and the corresponding maximum distance on January 6 is ~ 2500 km. This experimental fact may first imply that stronger atmospherics can be trapped by a magnetospheric path even if they are located at a great distance from the entrance region of the magnetospheric path. The attenuation rate of the Earth-ionosphere waveguide mode propagation at night is experimentally estimated to be ~ 4 dB/mm at frequency of 5 kHz [e.g., Wait, 1970], and so the total attenuation of 5-kHz wave is only

about 10 dB for a propagation distance of 2500 km from the thunderstorm to the ionospheric entrance point of whistler wave. Hence the propagation distance of the order of ~ 2500 km is not so influential, and the intensity of the causative atmospherics is of the predominant significance. Also, it is strongly indicative that such a preferred magnetospheric path is extremely rare in occurrence at these very low latitudes. Because if there could exist another magnetospheric path different from the ones on January 5 and 6 within the distance of 2500 km from the entrance point of the determined path, it would possibly trap the VLF wave energy from the atmospherics distributed widely within that distance, and it would be viewed from the station of ZJ on some occasions when its relative position with respect to the station is favourable for the viewing. However, in contradiction with this

expectation, other preferred paths have never been seen from our stations.

4. CONCLUDING REMARKS

The importance of the simultaneous location of the causative atmospherics and of the ionospheric exit regions of the whistlers has been demonstrated in this paper, and it will be of great potential in the detailed study of the propagation mechanism of whistlers along the total path from the source to the receiver.

REFERENCES

- Andrews, M. K., On whistlers with very low dispersion, *J. Atmos. Terr. Phys.*, **41**, 132, 1979.
- Bullough, K., and J. L. Sagredo, VLF goniometer observations at Halley Bay, Antarctica, I, The equipment and the measurement of signal bearing, *Planet. Space Sci.*, **21**, 899, 1973.
- Hayakawa, M., and Y. Tanaka, On the propagation of low-latitude whistlers, *Rev. Geophys.*, **16**, 111, 1978.
- Hayakawa, M., K. Ohta, and S. Shimakura, Spaced direction finding of nighttime whistlers at low and equatorial latitudes and their propagation mechanism, *J. Geophys. Res.*, **95**, 15,091, 1990.
- Leavitt, M. K., A frequency-tracking direction finder for whistlers and other very low frequency signals, *Tech. Rep. 3456-2*, Radiosci. Lab., Stanford Univ., Stanford, Calif., 1975.
- Liang, B. X., Z. T. Bao, and J. S. Xu, Propagation characteristics of nighttime whistlers in the region of equatorial anomaly, *J. Atmos. Terr. Phys.*, **47**, 999, 1985.
- Ohta, K., M. Hayakawa, and Y. Tanaka, Ducted propagation of daytime whistlers at low latitudes as deduced from the ground-based direction finding, *J. Geophys. Res.*, **89**, 7557, 1984.
- Okada, T., A. Iwai, and M. Hayakawa, The measurement of incident and azimuthal angles and the polarization of whistlers at low latitudes, *Planet. Space Sci.*, **25**, 233, 1977.
- Okada, T., A. Iwai, and M. Hayakawa, A new whistler direction finder, *J. Atmos. Terr. Phys.*, **43**, 679, 1981.
- Ondoh, T., M. Kohtaki, T. Murakami, S. Watanabe, and Y. Nakamura, Propagation characteristics of low-latitude whistlers, *J. Geophys. Res.*, **84**, 2097, 1979.
- Pierce, E. T., Monitoring of global thunderstorm activity, in *Planetary Electricity*, vol. 2, edited by S. C. Coroniti and J. Hughes, chap. 6, Gordon and Breach, New York, 1969.
- Shimakura, S., and M. Hayakawa, On the estimation of wave energy distribution of magnetospheric VLF waves at the ionospheric base with ground-based multiple electromagnetic field components, paper presented at Chapman Conference on Plasma Waves and Instabilities at Comets and in Magnetospheres, AGU, Sendai, Jpn., 1987.
- Shimakura, S., and M. Moriizumi, On the method of estimating the ionospheric height and propagation distance based on the dispersion effect of atmospherics, *Mem. Facult. Eng., Chiba Univ.*, **41**, 1, 1990.
- Shimakura, S., M. Moriizumi, and M. Hayakawa, Propagation mechanism of very unusual low-latitude whistlers with additional traces of the Earth-ionosphere waveguide propagation effect, *Planet. Space Sci.*, **39**, 611, 1991.
- Thomson, N. R., Ray-tracing the paths of very low latitude whistler-mode signals, *J. Atmos. Terr. Phys.*, **49**, 321, 1987.
- Wait, J. R., *Electromagnetic Waves in Stratified Media*, 2nd ed., p. 286, Pergamon, New York, 1970.
- Yoshino, T., Low-latitude whistlers and cloud distributions in the conjugate area, *J. Geophys. Res.*, **81**, 4793, 1976.
- M. Hayakawa, University of Electro-Communications, 1-5-5 Chofugaoka, Chofu, Tokyo 182, Japan.
- M. Moriizumi and S. Shimakura, Department of Electrical and Electronic Engineering, Chiba University, 33 Yayoi 1-chome, Chiba 260, Japan.
- K. Ohta, Department of Electronic Engineering, Chiba University, 1200 Matsumoto-cho, Kasugai, Aichi 487, Japan.

Trace splitting of whistlers: A signature of fine structure or mode splitting in magnetospheric ducts ?

D. Hamar¹, Cs. Ferencz¹, J. Lichtenberger², Gy. Tarcsai², A. J. Smith³ and K. H. Yearby⁴

(Received February 12, 1991; revised June 10, 1991; accepted August 12, 1991.)

Previously, we reported on the discovery of fine structure in whistler data received on the ground at Halley, Antarctica. This structure was not apparent in conventional spectral analysis but was revealed by the technique of digital matched filtering. We have now examined a larger data set, and a commonly observed phenomenon is that single whistler traces become split into two, over various frequency ranges. Examples are presented in the form of time-transformed spectrograms in which reference model whistlers are represented as vertical lines. The splitting is typically 5-15 ms (about 0.5% of the total whistler travel time) and extends over frequency ranges of a few hundred hertz which may occur anywhere between the upper and lower cutoff frequencies of the whistler. The splitting may be either symmetrical or unsymmetrical with respect to the unsplit trace. The effect is unlikely to arise in the spectrum of the lightning source or from propagation under or through the ionosphere. It may, however, be a signature of field-aligned fine spatial structure in plasmaspheric density, and hence refractive index, in the whistler duct. For simple longitudinal propagation, electron density fluctuations of the order of 1% and spatial scale sizes of the order of 50 km in the equatorial plane are implied. It seems possible that the observations could also be interpreted in terms of the mode theory of ducted propagation, assuming the excitation of two modes with group velocities differing by a few tenths of a percent.

1. INTRODUCTION

A matched filtering (MF) technique has been applied to broadband VLF data recorded digitally at Halley, Antarctica. Matched filtering produces dynamic spectra with higher resolution in both frequency and time than is possible using other methods. It has been proven earlier [Hamar *et al.* 1990]

that whistlers which appear as a single trace on a conventional spectrogram often contain many fine structure components. In the present study a whistler group was analyzed by MF. The group was generated by a multiple-stroke lightning flash and was composed of more than four smaller groups of traces belonging to different propagation paths. The splitting of individual whistler traces is tentatively interpreted as a signature of fine field-aligned structure in magnetospheric whistler ducts or by the existence of a number of guided modes.

¹Geophysical Research Group, Hungarian Academy of Sciences, Budapest, Hungary.

²Department of Geophysics, Eötvös University, Budapest, Hungary.

³British Antarctic Survey (NERC), Cambridge, United Kingdom.

⁴Department of Control Engineering, University of Sheffield, Sheffield, United Kingdom.

2. METHODS OF ANALYSIS

The matched filtering of whistlers and some of its results have been presented earlier [e.g., Hamar and Tarcsai, 1982]. It exploits the fact that the dispersion curve of a whistler component can be specified rather accurately in terms of just three parameters (which may, for example, be taken as D_0 , the zero frequency dispersion, f_n , the nose frequency, and t_0 , the

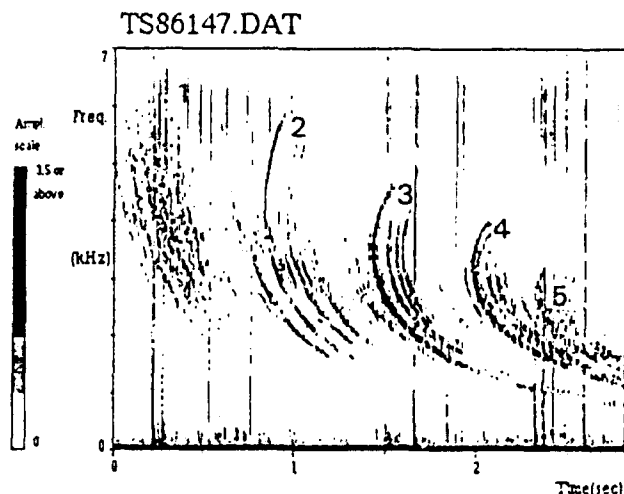


Fig. 1. A conventional FFT spectrogram of a whistler group recorded at Halley on August 10, 1986, at 1452:09 UT. The subgroups labeled 1-5 are associated with different paths through the magnetosphere.

epoch of the whistler source). MF is a correlation between the measured whistler signal and a model theoretical wave form constructed from the theoretical whistler dispersion relation together with parameters which give the best fit to the data. For a whistler trace it yields the time of arrival and the approximate amplitude of a given frequency. Any signals or noise which do not approximately obey the model whistler dispersion are strongly discriminated against.

The filter output is smoothed and the time resolution obtainable is some fraction of $2/\Delta f$ anywhere along the whistler, regardless of df/dt ; this is a great improvement on

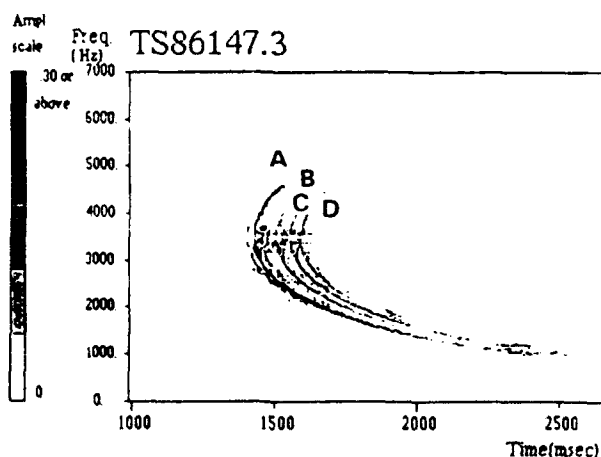


Fig. 2. Group 3 shown after MF analysis. The four traces labeled A-D correspond to four separate lightning discharges. Noise has been suppressed.

the conventional Fourier analysis technique. In practice, a filter bandwidth of $\Delta f = 200$ Hz has been used, which proved to be optimum in the analysis of Antarctic whistlers. The method enables one to obtain an accuracy of the order of 0.1 ms in whistler travel times and to separate two closely spaced components provided that they are separated by at least ≈ 3 ms.

Because of the MF analysis, at such extreme resolution, is becoming somewhat sensitive to the accurate description of the signal form and to the details of the applied mathematical procedures, the signal analysis should be accompanied by a rigorous analysis of the MF procedure itself and also that of the theory of the whistler mode propagation. As our model computations and other considerations demonstrated, the effects and results presented in this paper are genuine. However, the details of the splitting can be even more accurately studied by a parameter estimation procedure applied to the MF output and by a more careful treatment of the

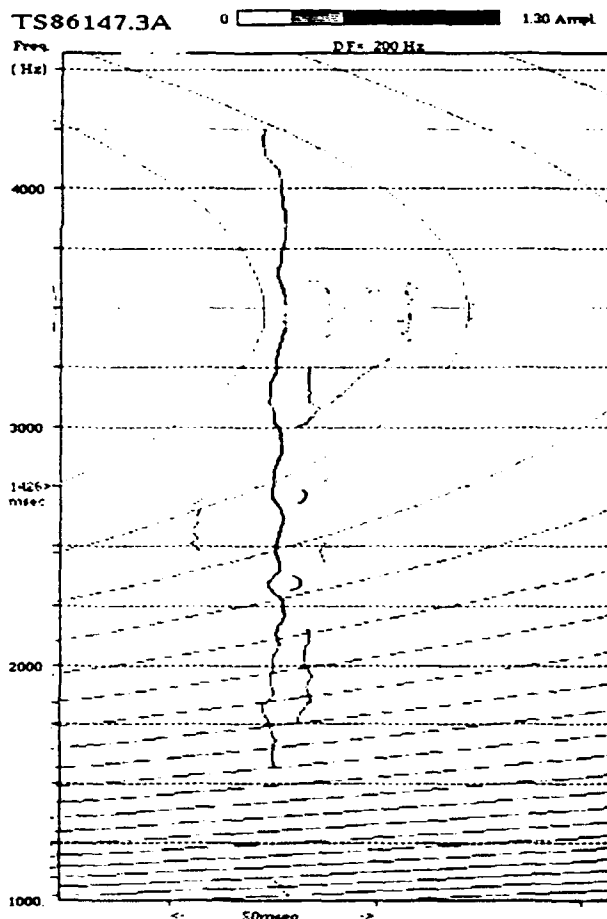


Fig. 3. A transformed and expanded spectrogram showing trace A of group 3. The dashed hyperboloid lines are isochrones.

spurious signals (computational artifacts) which might appear around the nose frequency. These questions, together with further results, will be presented in another paper.

3. EXPERIMENTAL DATA

For the purposes of this paper, we study a whistler group recorded at Halley, Antarctica (76°S, 27°W, $L \approx 4.3$), on August 10, 1986, beginning at 1452:09 UT, using two vertical loop aerials aligned north-south and east-west. The sampling rate was 14.18 kHz per component, and the length of sample was 2.87 s. The result of conventional (fast Fourier transform (FFT)) spectral analysis is shown in the spectrogram of Figure 1. The signals present are typical multiple-path whistlers, generated by a multiple-stroke lightning discharge. As further analysis showed, there were at least four clearly discernible strokes in the lightning flash, separated by 37–78 ms. Thus each whistler subgroup, denoted by numbers 1 to 5 and propagating along various L shells, is composed of at least four traces, generated by the successive strokes. The various traces in groups 2–5 were scaled for f - t values, first from the FFT spectrograms and then by using the matched filter technique, and then analyzed by an accurate curve fitting procedure.

Group 3 is particularly remarkable because all four traces in this group followed extremely closely the same propagation path (within an L range of 0.001, i.e., 6 km(!) in the equatorial plane). Therefore in the present paper we focus our attention on this group. The characteristic parameters obtained by curve fitting for group 3 whistlers are the following:

| | |
|-----------------------------|---|
| equatorial gyrofrequency | $f_{Heq} = 9176 \pm 4$ Hz; |
| nose frequency | $f_n = 3408 \pm 2$ Hz; |
| dispersion | $D_0 = 92.2 \pm 0.1$ s ^{1/2} ; |
| L value | $L = 4.566 \pm 0.001$; |
| equatorial electron density | $n_{eq} = 219 \pm 1$ cm ⁻³ ; |
| tube electron content | $N_T = (3.52 \pm 0.01) \cdot 10^{13}$ /cm ² -tube. |

The first three parameters (or at the beginning of the analysis their less accurate values) were used in constructing the theoretical whistler waveform needed in matched filtering. The first, strongest, trace is followed by three weaker traces of the same dispersion.

4. FINE ANALYSIS

In the MF analysis the epochs and amplitudes of the maxima of smoothed MF output envelopes were determined at each frequency. The MF was normalized to obtain constant energy per unit bandwidth at each frequency. The filter length and therefore the noise suppression of MF depends on the frequency. It is lowest at the whistler nose and increases toward higher and lower frequencies.

The MF was performed at steps of 5 Hz between 1 and 5 kHz. The (f, t) pairs and amplitudes, obtained from the peaks of the smoothed envelopes of the filter outputs, give a dynamic spectrum (Figure 2) with the noise strongly suppressed. The main trace (A) and three other signals (B, C, D) following A by 78, 126, and 163 ms respectively, are clearly seen in this figure.

The higher time resolution of MF is better seen by "straightening" the whistler trace, i.e., by subtracting the theoretical (fitted) travel times from the measured $t(f)$ data. A model whistler is represented by a straight vertical line on such transformed plots, while spherics would appear hyperboloid, like the isochrones in Figure 3 which shows a transformed plot of trace A.

5. TRAVEL TIME RESIDUALS AND TRACE SPLITTING

As it can be seen for trace A, the transformed traces locally deviate more or less from the vertical; these deviations are known as travel time residuals [Tarcsei *et al.*, 1989; Hamar *et al.*, 1990]. Figure 4 shows the curves of travel time residuals for traces A, B, and C. It is seen that the residual curves are very similar to each other, i.e., certain frequency ranges appear to travel faster (e.g., at 1.2, 1.6, 1.8, 2.3 kHz) others slower (e.g., at 1.4, 2.2, 2.4 kHz, etc.) than the theoretical signal. Such strong similarity in residuals occurs for other groups too (e.g., group 4), covering, however, narrower frequency ranges than at group 3 whistlers.

Furthermore, it is seen that at certain frequencies, trace A is split into two signals. The same holds for signals B, C, and D (Figure 5), which are strikingly similar to trace A. The

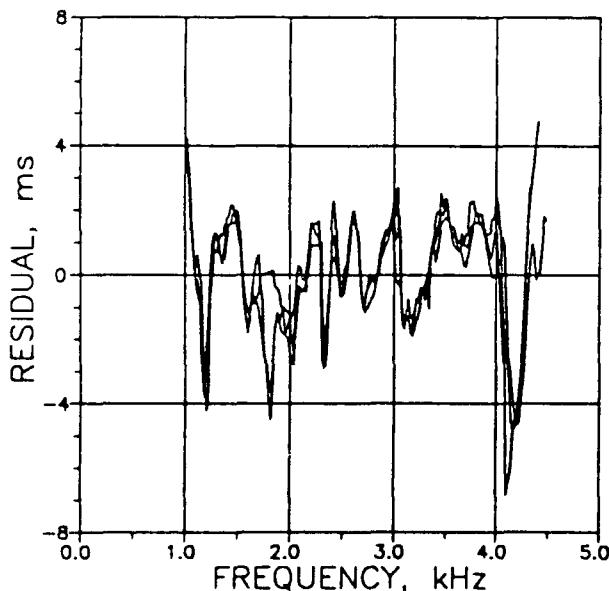


Fig. 4. Travel time residuals of group 3, traces A, B, and C.

splitting covers a frequency range of 80 Hz at 2.3 and 2.7 kHz, and 400 - 500 Hz at 1.76 to 2.16 kHz and 3.0 to 3.5 kHz. Thus there are several cases of traces split over a given frequency range, either quasi-symmetrically (e.g., at 2.3 kHz) or asymmetrically (e.g., around 2 kHz) with respect to a vertical theoretical trace. Sometimes a decrease in the amplitude of the split traces, as compared to that of the unsplit trace, is also evident. The split parts are separated by amounts varying between 5 and 15 ms, apparently not depending on the frequency. Similar splitting or fine structure has been found in other whistler components shown in Figure 1 (see Figure 6) and in other data sets. The same few (three or four) splitting patterns were observed in the majority of cases.

6. DISCUSSION

For the purposes of discussion, we consider an idealized localized splitting event illustrated in Figure 7. The

characteristics shown are typical of those found in the data, which have been described above. Let us now consider possible origins for a splitting of $\delta t = 8$ ms. It is conceivable that the splitting could be present in the source spectrum of the lightning; however there is no evidence that lightning displays such features, and the fact that the effect is persistent for several distinct flashes makes this possibility seem unlikely. Thus the origin of the splitting would appear to arise in the course of propagation of the signal from the lightning source to the receiver. The total group travel time t_g is composed of three parts: subionospheric, transionospheric, and magnetospheric (t_m) components.

For the example event of Figure 7 these parts are approximately 10 ms, 100 ms, and 2 s respectively. Any subionospheric effect would thus be too small to produce a δt of 8 ms. It is possible that the splitting could occur during the propagation of the whistler through the ionosphere, but rather large variations in group path, of the order of 8%, would be

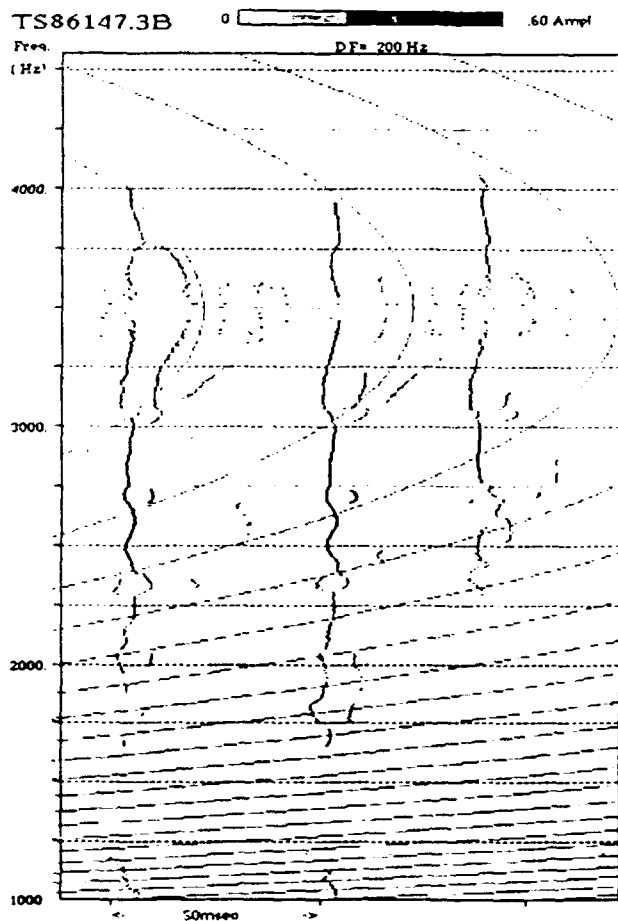


Fig. 5. A transformed and expanded spectrogram showing traces B, C, and D of group 3.

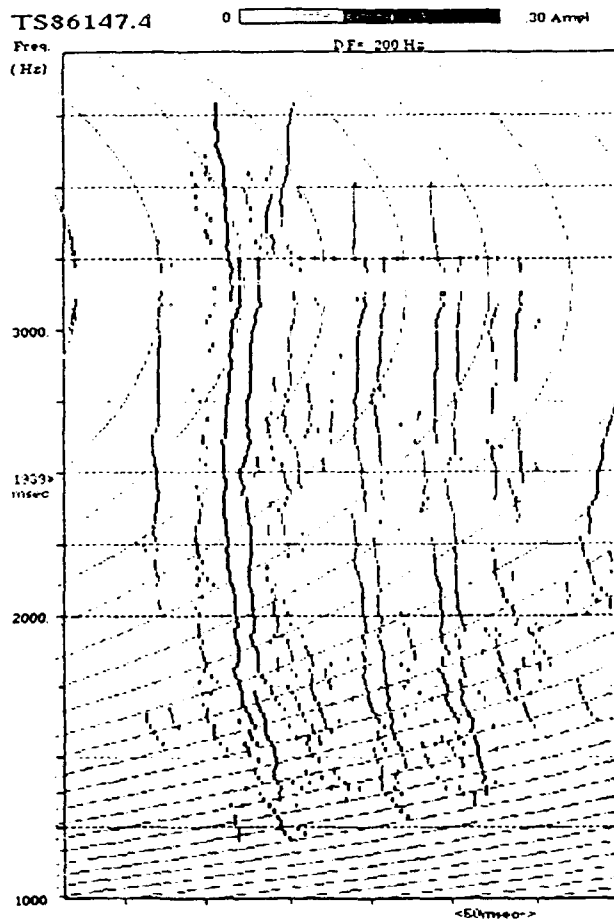


Fig. 6. The transformed spectrogram of group 4 whistlers. Traces a, b, and c correspond to traces A, B, and C of group 3, respectively. The splitting covers wide frequency bands. The dashed hyperboloid lines are isochrones.

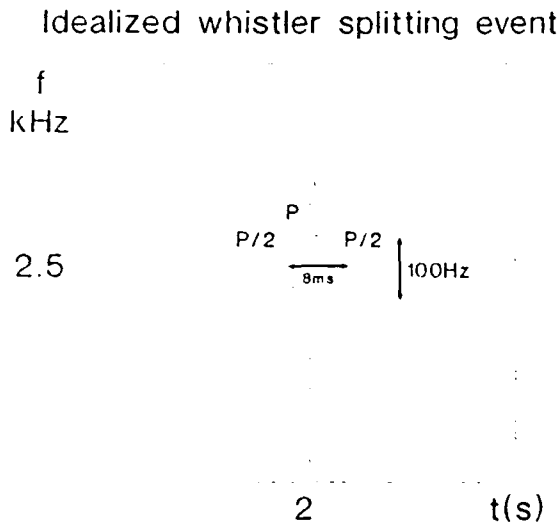


Fig. 7. An idealized whistler trace splitting event. The time splitting is $\delta t = 8$ ms, over a frequency range $\Delta f = 100$ Hz. The split is centered at $f = 2.5$ kHz and $t = 2$ s measured from the causative spheric. (NB. This is in contrast to the spectrograms shown in Figures 1-3 and 5, for which the time origin is the start of digital data frame and does not coincide with the causative spheric.)

required; to achieve this, large gradients in ionospheric refractive index would need to be postulated. In addition the character of splitting (Figures 3, 5, and 6) can hardly be explained by any conceivable subionospheric or transionospheric effect. Considering that the splitting seems to be a commonly occurring feature, it seems more likely that it represents fine structures on the magnetospheric part of the path, for which it represents 0.4% of t_m , or mode splitting.

If this is indeed the case, then a proper explanation of the effect should rely on a theory of the ducted propagation of whistlers as applied to the case where fine structure exists in the ducts. The approach might be that of ray tracing [e.g., *Strangeways*, 1982], mode theory [e.g., *Scarabucci and Smith*, 1971; *Laird and Nunn*, 1975], or an analytical theory such as that of *Karpman and Kaufman* [1982]. Such treatment is beyond the scope of the present paper.

If the trace splitting is attributed to fine structure in the magnetospheric ducts then we can attempt to estimate the order of magnitude of the density fluctuations and scale sizes involved, by assuming simple longitudinal propagation. In this case a structure in group time of the order of $(\delta t/t_m) \sim 0.4\%$ implies a corresponding structure in equatorial electron density of $(\delta n_{eq}/n_{eq}) \sim 0.8\%$ and in cross- L scale size of $(\delta L/L) \sim 0.2\%$ (assuming $n_{eq} \sim L^{-2}$). At $L = 4.5$ the latter correspond to ~ 60 km in the equatorial plane and ~ 4 km at the ionosphere. The figures are about an order of magnitude smaller than those usually considered to apply to whistler ducts as a whole, i.e., those structures responsible for the coarse level of discrete

traces in Figure 1, and are therefore quite feasible as a signature of fine structure within such ducts. It may well occur, that the roughly parallel traces, separated by 15-20 ms and extending over a range of 1-2 kHz, as observed for certain group 2, 3 and 4 whistlers (or for group 4 in Figure 1), can be actually interpreted in terms of fine structures within the ducts. In a strict sense, however, we do not consider such traces as splitted ones. It is also conceivable that such closely spaced traces appear as a result of multimode and coupled multimode propagation in the waveguide structure.

We may also note at this point that the figure of a few tenths of a percent in the group time splitting is of the same order as the fractional difference between the group velocities of different modes calculated by *Scarabucci and Smith* [1971] for bell-shaped ducts. It is to be noted, however, that for the homogeneous ducts used by *Scarabucci and Smith* [1971] the splitting appears at high frequencies, mainly above the nose frequency. Although we frequently observed splitting around and above the nose, the phenomenon often appears well below the nose too. Therefore it is clear that before drawing further conclusions the mode theory ought to be extended for inhomogeneous (variable along the magnetic field line) ducts. It is not at present clear why the observed splitting occurs only over a limited frequency range.

With respect to the residual curves it is important to note that the largest similarity, covering a frequency range of 350 Hz, occurs for group 3 whistlers, which follow very closely the same path (see their L value and n_{eq} Figure 8). The spread in L and equatorial electron density is larger for group 4 for which similar residuals were observed only for 300-600 Hz wide bands and even larger for group 2 and 5, where the resemblance between the residual curves is limited to even narrower frequency bands. It is clear that the similarity between the fine details of residual curves of different traces

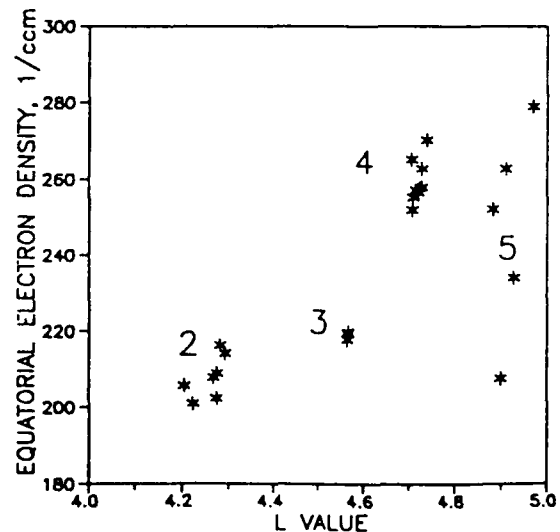


Fig. 8. Equatorial electron density versus the L values of the propagation computed from group 2 to 5 whistlers.

demonstrates their reliability and yields information on the propagation conditions. As it has been shown by Strangeways [1978] and stressed by Tarcsai *et al.* [1989], the residual patterns for whistlers trapped in fine structure ducts will be more complicated than those trapped in ducts with simple structure. Therefore the further analysis of residuals may help in gaining more information on the fine structure of ducts and thus in judging about the cause of trace splitting too. More realistic model calculations, for comparison with the experimental data, are required to test the above suggestions.

7. CONCLUSIONS

1. Digital matched filtering of digitally recorded ground-observed whistlers is revealing new features of their fine spectral structure.
2. It is believed that this fine structure may be a signature of structure in magnetospheric field-aligned irregularities ("whistler ducts").
3. A commonly observed feature is the splitting of whistler spectra ($\delta t \sim 5$ -15 ms) over a limited or wide frequency range.
4. This can occur anywhere in the whistler bandwidth.
5. For a simple longitudinal propagation model we infer density fluctuations of a few tenths of a percent and spatial (cross- L) scales of tens of kilometers in the equatorial plane.
6. The observations may be explicable in terms of the excitation and propagation of different guided modes with differing group velocities.
7. More detailed calculations, with realistic models, are required to test these interpretations of the experimental results and to exhaust the research opportunities provided by high resolution digital matched filtering.

Acknowledgment. This research was supported in part by the National Fund for Scientific Research (OTKA, Hungary) through grant 1292/86.

REFERENCES

- Hamar, D. and Gy. Tarcsai, High resolution frequency-time analysis using digital matched filtering, Part I, Theory and simulation study, *Ann. Geophys.* **38**, 119-128, 1982.
- Hamar, D., Gy. Tarcsai, J. Lichtenberger, A.J. Smith, and K.H. Yearby, Fine structure of whistlers recorded digitally at Halley, Antarctica, *J. Atmos. Terr. Phys.*, **52**, 801-810, 1990.
- Karpman, V.I., and R.N. Kaufman, Whistler propagation in ducts, *J. Plasma Phys.*, **27**, 225-238, 1982.
- Laird, M.J., and D. Nunn, Full-wave VLF modes in a cylindrically symmetric enhancement of plasma density, *Planet. Space Sci.*, **23**, 1649-1657, 1975.
- Scarabucci, R.R., and R.L. Smith, Study of magnetospheric field oriented irregularities - The mode theory of bell-shaped ducts, *Radio Sci.*, **6**, 65-68, 1971.
- Strangeways, H.J., An investigation of the propagation of whistlers in magnetospheric ducts by means of ray-tracing, curve fitting and direction-finding techniques, Ph.D. thesis, Southampton Univ., Southampton, 1978.
- Strangeways, H.J., The effect of multiduct structure on whistler-mode wave propagation, *J. Atmos. Terr. Phys.*, **44**, 901-912, 1982.
- Tarcsai Gy., H.J. Strangeways, and M.J. Rycroft J. [1989], Error sources and travel time residuals in plasmaspheric whistler interpretation, *J. Atmos. Terr. Phys.*, **51**, 249-258, 1989.
- Cs. Ferencz and D. Hamar, Geophysical Research Group, Hungarian Academy of Sciences, Budapest, Hungary.
- J. Lichtenberger and Gy. Tarcsai, Department of Geophysics, Eötvös University, H-1083 Budapest, Kun B.t. 2, Hungary.
- A. J. Smith, British Antarctic Survey (NERC), Madingley Road, Cambridge, United Kingdom.
- K. H. Yearby, Department of Control Engineering, University of Sheffield, Sheffield, United Kingdom.

Ion cyclotron emissions observed by the satellite Akebono in the vicinity of the magnetic equator

Yoshiya Kasahara, Akira Sawada, Masayuki Yamamoto, and Iwane Kimura
Department of Electrical Engineering, Kyoto University, Kyoto, Japan

Susumu Kokubun and Kanji Hayashi
Institute of Geophysics, University of Tokyo, Tokyo, Japan

(Received March 11, 1991; revised July 1, 1991; accepted July 8, 1991.)

In our wave observations by a wave measuring instrument named ELF on board the Akebono satellite, interesting ELF emissions in the frequency ranges above the cyclotron frequencies of H^+ and/or O^+ ions have been detected. This paper is to propose that these emissions are due to electromagnetic ion cyclotron waves, which might have been generated by the ion cyclotron resonant instability due to a temperature anisotropy of hot H^+ ions, trapped along the magnetic field lines around the geomagnetic equatorial plane. In order to confirm the characteristics of these ELF emissions, statistical characteristics, refractive index, and wave normal direction of the emissions are first estimated. The propagation characteristics of these ELF emissions are examined by ray tracing for ion cyclotron waves in the magnetosphere using both cold and hot plasma models.

1. INTRODUCTION

By the GEOS 1 and 2 satellites, measurements of ULF-ELF waves were made near the plasmapause and ULF-ELF emissions in a frequency range near the local cyclotron frequencies of H^+ and He^+ ions (0.4 ~ 2 Hz) were observed around the geomagnetic equatorial region. Similar emissions were also observed by the ATS-6 satellite. They were interpreted as ion cyclotron waves (ICW) associated with the existence of He^+ ions [Young et al., 1981; Mauk et al., 1981; Roux et al., 1982]. The propagation characteristics of these ELF emissions were confirmed by ray tracing in a cold plasma model [Rauch and Roux, 1982]. Several studies have been made for the wave-particle interaction between ICW and multicomponent ions by using analytical results and simulation techniques [Gendrin et al., 1984; Omura et al., 1985; Gomberoff and Vega, 1987].

The Akebono satellite included a wave measuring instrument named ELF covering a frequency range from a few hertz to 80 Hz for one component of E and three components of B fields measured by a 60-m tip-to-tip wire antenna and three orthogonal search coil antennas [Kimura et al., 1990].

In our observations of ELF emissions in the vicinity of the geomagnetic equatorial region it was found that the emissions are very similar to those observed by the GEOS satellite, although the observed location of the phenomena is greatly different. The frequency range of the ELF instrument covers the local cyclotron frequencies of H^+ , He^+ , and O^+ ions, and most

of the wave phenomena are considered to be associated with ion constituents.

The present paper discusses these ELF emissions observed by the Akebono satellite in a manner similar to that of Young et al. [1981] and Roux et al. [1982]. An outline of the present paper is as follows. In the next section the results of data analyses of the ELF waves observed by the Akebono satellite are presented. In section 3, with reference to the generation mechanism of the emissions, we will present the results of three-dimensional ray tracing for ion mode waves using both a cold and a hot plasma model. Propagation characteristics, spatial distribution, and growth rates of the emissions are clarified by our ray tracing. Finally, a summary and discussions are added in the last section.

2. ELF EMISSIONS OBSERVED AROUND THE GEOMAGNETIC EQUATORIAL REGION

2.1. General characteristics of the ELF emissions

Figure 1 shows an example of the ELF emissions observed from 0342 to 0358 UT on May 27, 1989, when the Akebono satellite was tracked from Kagoshima Space Center (KSC) in Japan. ALT, MLT, and MLAT stand for altitude (kilometers), magnetic local time, and magnetic latitude, respectively, of the trajectory of the Akebono satellite. The vertical dashed line indicates the location of the geomagnetic equator. The Akebono satellite was traversing from the southern hemisphere to the northern hemisphere. The telemetry reception was terminated on this orbit nearly at the geomagnetic equator for the sake of ranging the Akebono satellite.

In this figure the bands of emissions were observed for the geomagnetic latitude range from -28° to nearly 0° . It is clearly

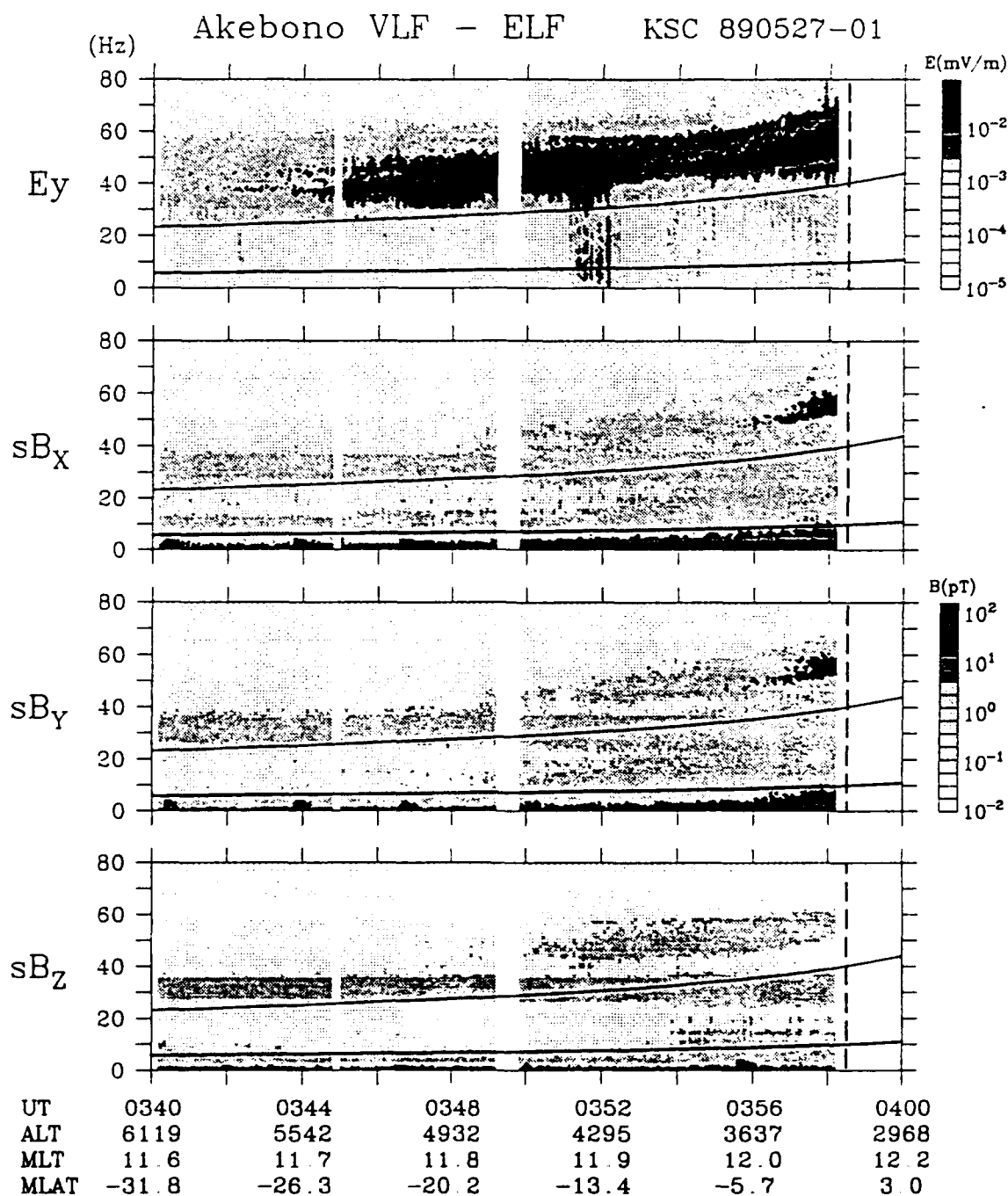


Fig. 1. A f - t diagram of the ELF emissions observed on May 27, 1989. The vertical dashed line indicates the location of the geomagnetic equator. The upper solid line indicates the local He^+ cyclotron frequency, and the lower solid line indicates the local O^+ cyclotron frequency.

Akebono VLF - ELF KSC 890527-01

TIME = 3:58:00.480(UT) f_{He^+} = 156.6(Hz)
 ALT = 3302.7(km) f_{He^+} = 39.2(Hz)
 MLT = 12.1 f_{O^+} = 9.6(Hz)
 MLAT = -1.6

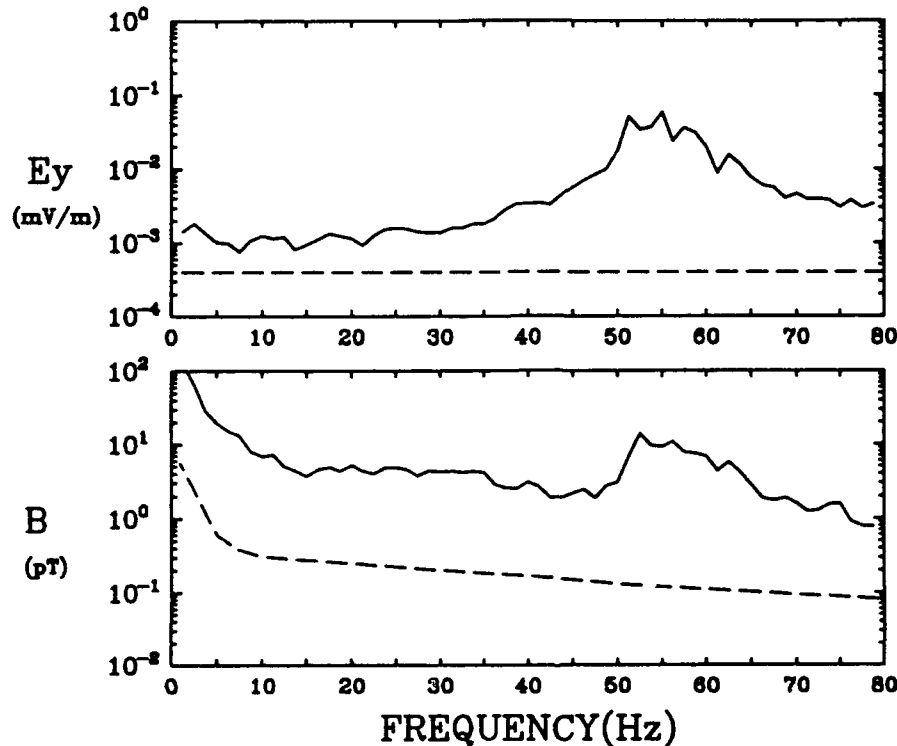


Fig. 2. Frequency dependence of the electric and magnetic intensity of the ELF emissions observed on May 27, 1989. The dashed lines represent the threshold levels for electric field and magnetic field channels respectively.

seen that the frequencies of the emissions are always above the He^+ cyclotron frequency (f_{cHe^+}) which is shown by the upper solid line in the figure, and there is a sharp gap between the bottom of the emissions and the f_{cHe^+} . It is distinctive that these emissions have magnetic field components, so that these are identified as one of the electromagnetic ion cyclotron modes. A dark horizontal band around 32 Hz, most apparent in the magnetic field components, is caused by interference.

Figure 2 illustrates the frequency dependence at 0358:00 UT of E_y and B after an fast Fourier transform (FFT) analysis, where

$$B = \sqrt{sB_x^2 + sB_y^2 + sB_z^2}. \quad (1)$$

The frequency resolution is 1.25 Hz, and dashed lines in the figure represent the threshold levels for electric field and magnetic field channels respectively. The internal interference around 32 Hz, seen in Figure 1, has been removed in Figure 2. From the figure the maximum electric and magnetic fields are 60 μ V/m and 14 pT, respectively.

Figure 3 shows an event observed from 1034 to 1047 UT on July 14, 1989, in which two bands of emissions coexist, one above f_{cHe^+} (the upper solid line in the figure) and the other one above O^+ ion cyclotron frequency (f_{cO^+}) (the lower solid line). It is to be noted that the geomagnetic latitude range where the lower band emissions were observed is wider than the range of the upper band emissions.

In this figure, because the intensities of the electric components are very weak in comparison with the case on May 27, 1989 (the peak intensities of both emissions are about 1.5 μ V/m), the magnetic components were under the threshold noise level and were not identified. In some cases with two emission bands coexisting, however, the magnetic field component is also identified. Therefore these emissions are considered to be one of the electromagnetic ion cyclotron modes.

These examples indicate that the frequencies of these waves are dependent on the local cyclotron frequencies of ions, that is, they are closely connected with the presence of He^+ and O^+ ions in addition to H^+ ions in the magnetospheric plasma. In this

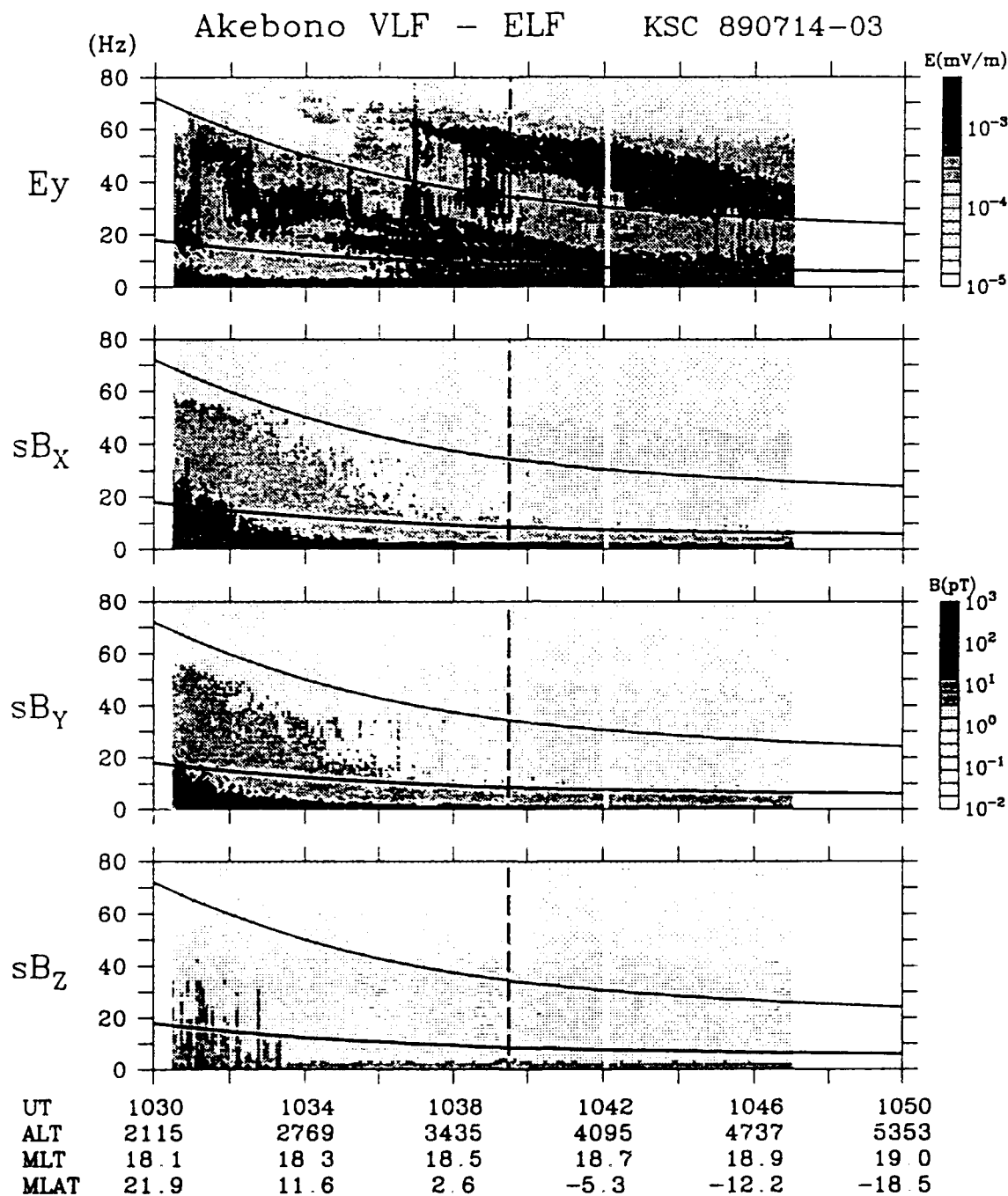


Fig. 3. A f - t diagram of the ELF emissions observed on July 14, 1989. The vertical dashed line indicates the location of the geomagnetic equator. The upper solid line indicates the local He^+ cyclotron frequency and the lower solid line indicates the local O^+ cyclotron frequency.

paper we propose that these emissions are electromagnetic ion cyclotron waves (ICW) trapped along the magnetic field lines around the geomagnetic equatorial region. The details of the ICW will be discussed in the following sections. The generation mechanism of these emissions will be discussed in section 3. In the following we will call the upper band of emissions above f_{cHe^+} the H^+ mode, and the lower band of emissions above f_{cO^+} the He^+ mode.

2.2. Statistics of the ELF emissions

In this section we will show the results of the statistical analyses of the ELF emissions we observed. In general, the characteristics of the H^+ mode emissions are very similar to those of the emissions observed by the GEOS 1 and 2 satellite [Young et al., 1981], though the absolute frequencies observed by us were much higher than those observed by the GEOS 1 and 2, since the locations of our observations are much lower than the GEOS observations. On the other hand, there have been no previous reports of emissions similar to the He^+ mode emissions presented in this paper.

2.2.1. Spatial and time distribution of the ELF emissions.

We have analyzed the data of 775 orbits tracked at KSC, Prince Albert (PA) in Canada and Esrange (ESR) in Sweden from April 1989 to November 1990, although we have not yet completed the examination of all the data in this term. In the analyses we have found 81 H^+ mode emissions and 48 He^+ mode emissions, in which we do not count emissions whose mode is not clearly identified. For 48 passes in which the He^+ mode emissions were observed the H^+ mode emissions were simultaneously observed.

The trajectories of the Akebono satellite while the H^+ mode emissions were observed are shown in Figure 4, which includes the cases where the H^+ mode and He^+ mode emissions were simultaneously observed. On the other hand, the trajectories for the He^+ mode emissions are shown in Figure 5. In Figures 4 and 5 the left panels show the projection of the trajectories on the geomagnetic meridian plane and the right panels show their projection on the geomagnetic latitude-longitude plane.

In these figures we can find that most of the H^+ mode emissions were observed in a region lower than 20° in geomagnetic latitude. It is interesting to note that the observed region of the He^+ mode emissions is wider in latitudes than the region of the

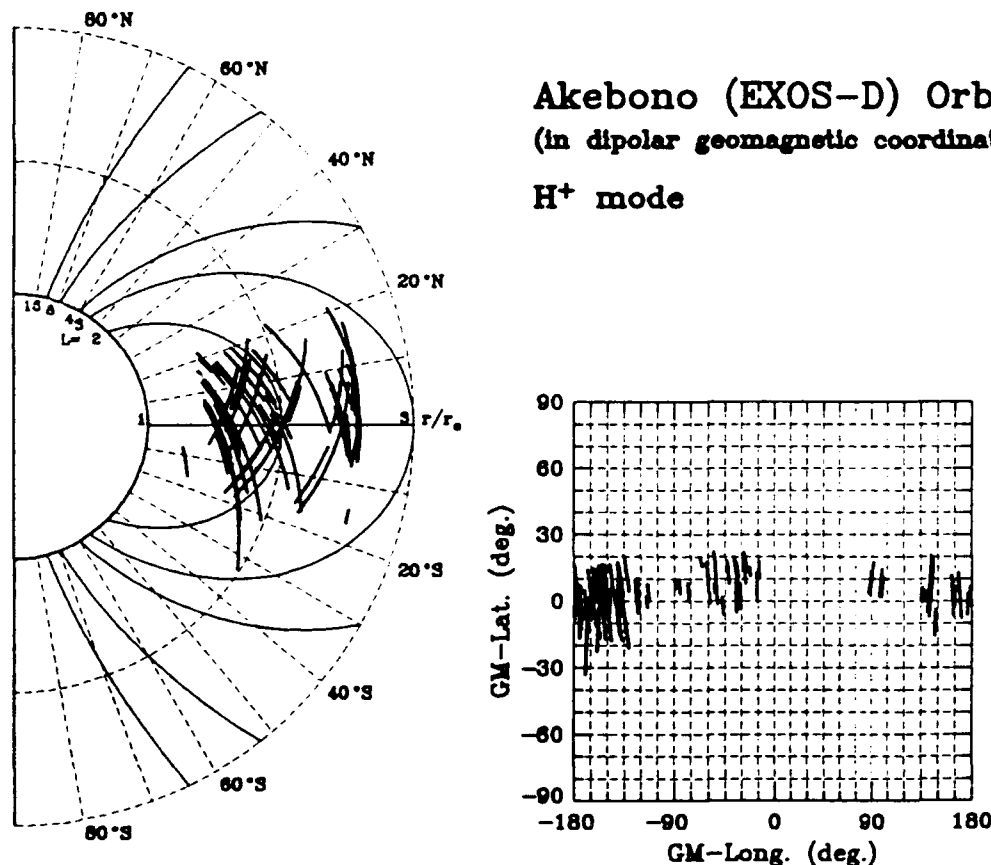


Fig. 4. Trajectories of the Akebono satellite where the H^+ mode emissions were observed.

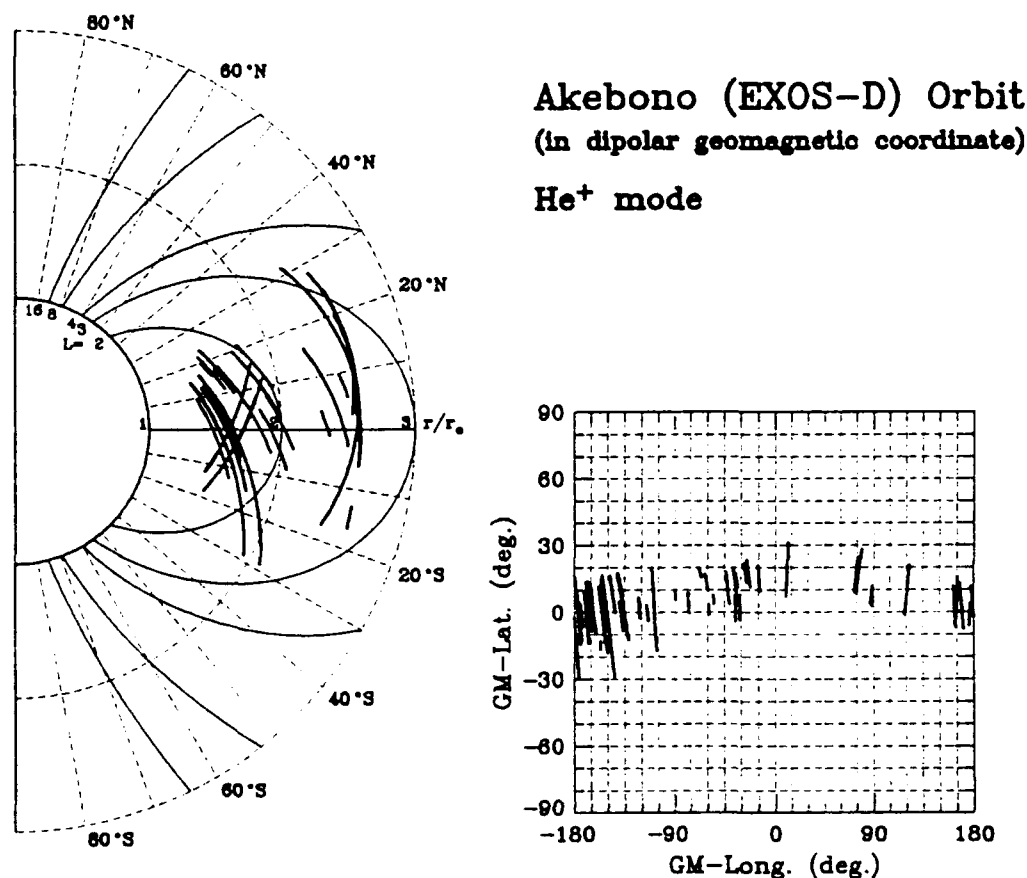


Fig. 5. Trajectories of the Akebono satellite where the He^+ mode emissions were observed.

H^+ mode emissions. The trajectories of the Akebono satellite range in altitudes from the perigee of about 300 km to the apogee of about 10,000 km in these 2 years, and it is remarkable that both the H^+ mode and He^+ mode emissions were observed even at altitudes as high as 10,000 km (nearly the apogee altitude). These observations suggest that He^+ and O^+ ions exist at such high altitudes.

Figure 6 is a histogram of the occurrence number of the ELF emissions versus altitude. The vertical axis indicates the number of trajectories when the ELF instrument was operating around the geomagnetic equator. The horizontal axis indicates the altitude of the Akebono satellite passing through the geomagnetic equator. The cross-hatched region indicates the number of trajectories when the two bands emissions of H^+ mode and He^+ mode coexist. The single-hatched region indicates those in which one band emissions, that is, only the H^+ mode emissions were observed during the trajectory. The blank region indicates none of the ELF emissions was observed around the geomagnetic equator even though the ELF instrument was in operation.

As shown in Figure 6, the H^+ mode emissions were observed very frequently in the altitude range from 3000 to 7000 km. Es-

pecially in the altitude range from 3000 to 5500 km, the two-band emissions, that is, H^+ mode and He^+ mode coexist with a very high occurrence probability. In the higher-altitude range around 10,000 km, the occurrence probabilities of the emissions decrease, but some emissions were observed because there are very many orbits in such a high-altitude range. The one band, that is, only H^+ mode emissions are likely to occur more frequently than the two-band emissions.

In the same way, Figure 7 is a histogram of occurrence number of the ELF emissions versus the magnetic local time of the trajectories. The horizontal axis indicates the magnetic local time when the Akebono satellite has passed through the geomagnetic equator. Both modes (especially the He^+ mode) emissions were often observed in the eveningside. On the other hand, the H^+ mode emissions were also observed in the morning to the dayside. Since the observational coverage from the nightside to the morningside are insufficient, occurrence characteristics in such a region are not clear at present.

2.2.2. Relationship with K_p index. We have surveyed a relationship between the K_p index and the ELF emissions. So far, K_p indices until August 31, 1990, were available.

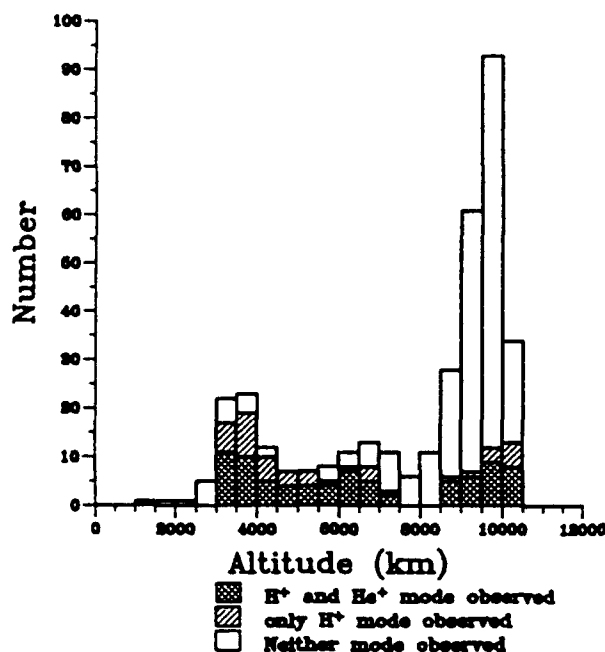


Fig. 6. Altitude dependence of the occurrence probabilities of the H^+ mode and the He^+ mode emissions.

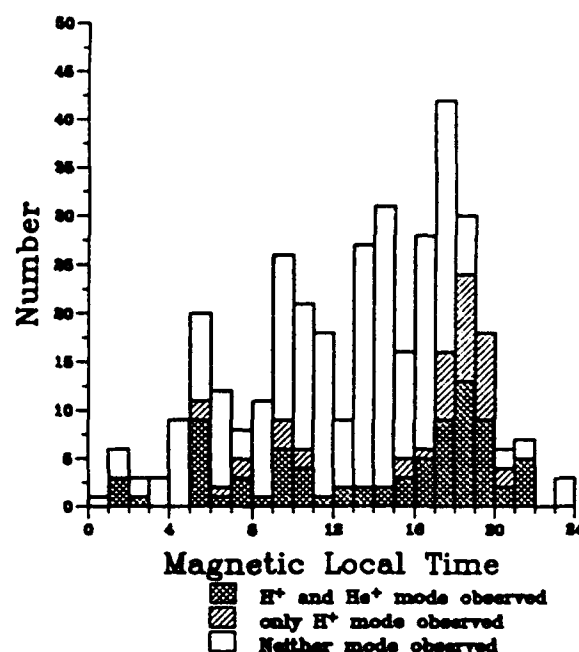


Fig. 7. Magnetic local time dependence of the occurrence probabilities of the H^+ mode and the He^+ mode emissions.

Figure 8 is a histogram which indicates this relationship. The vertical axis is the same as in Figures 6 and 7 and the horizontal axis indicates the K_p index when the Akebono satellite passed through the geomagnetic equator. It is evident from this figure that when the geomagnetic activity was high, these emissions were observed with a high occurrence probability. However, even during times when the geomagnetic activity was very quiet these emissions were also observed. There is no clear difference in the relationship with magnetic activity between the one-band case and the two-band case.

2.3. Refractive index and wave normal direction

2.3.1. *Refractive index.* The refractive index of each mode was roughly estimated, by using the observed one component of electric field and three components of magnetic field, i.e., by the following equation:

$$\mu \simeq \frac{cB}{E_i} \quad (i = x \text{ or } y), \quad (2)$$

where c is the light velocity and B is calculated by (1). Actually, however, when the angle between the geomagnetic field line and the wave normal direction is nearly perpendicular, the refractive index cannot be estimated correctly by this equation.

In the case of Figure 2 observed on May 27, 1989, the refractive index calculated by (2) is of the order of 100 in the frequency range from 45 to 60 Hz. In general, the quantity of the refractive index of the H^+ mode determined by (2) at the altitude

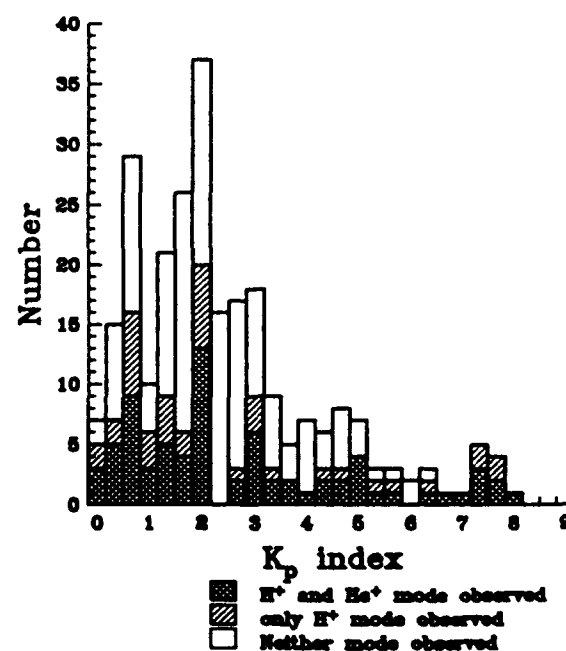


Fig. 8. Relation between the K_p index and the occurrence probabilities of the H^+ mode and the He^+ mode emissions.

range from 3000 to 4000 km is from 100 to 200, and the one of the He^+ mode is 400 to 600.

2.3.2. Determination of wave normal direction. The study of the wave normal vector is important to identify the mode of the waves and to interpret the propagation characteristics of them. In the present study we have three components of wave magnetic fields obtained by search coils. By using these wave magnetic field data the wave normal direction can be determined.

At first, wave magnetic field data of the ELF emissions are frequency-analyzed by FFT with a band width of 1.25 Hz and spectral matrix elements are calculated from these frequency components. The eigenvalues of the three-dimensional spectral matrix are determined as λ_m (where $m = 1, 2, 3$), and corresponding eigenvectors as u_m . If there are not more than two nonzero eigenvalues, the number of the nonzero eigenvalues corresponds to the number of the waves, and the wave normal direction k_m for each eigenvector is given by

$$k_m \parallel \text{Re}(u_m) \times \text{Im}(u_m), \quad (3)$$

where the double slash indicates that these two vectors are parallel to each other and "Re" and "Im" indicate real and imaginary part of the eigenvector, respectively [Samson and Olson, 1980]. If there are three eigenvalues, there are more than three waves and we cannot determine their wave normal directions.

After the FFT analyses we extract components with a bandwidth of 6.25 Hz from the ELF emissions, which correspond to five frequency components of 1.25-Hz resolution around the frequency of the maximum signal intensity.

We then calculate the spectral matrix elements and eigenvalues. When the ratio of the smallest eigenvalue to the largest one exceeds the threshold level ($= 0.01$), we assumed that there are more than three waves, and when the ratio of the second largest to the largest exceeds the threshold level, there are two waves.

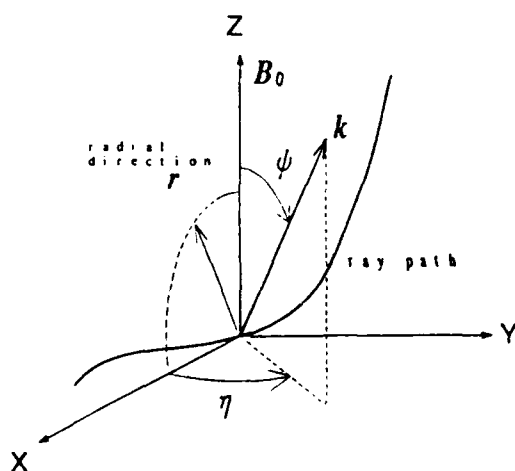


Fig. 9. Coordinate system for representing the wave normal vector.

In the case of the H^+ mode waves observed from 0350:00 to 0358:20 UT on May 27, 1989 (see Figure 1), 5.9% of all the spectral matrices have only one eigenvalue, 74.8% and 19.2% of them have two and three eigenvalues, respectively. Especially in the case of two eigenvalues, 36.2% of them are nearly the same order of eigenvalues. This fact implies that multiple waves usually coexist around the geomagnetic equatorial region.

Although in case there are multiple waves coexisting, it has been known that the determination of the wave normal direction by using the eigenvectors are not always correct, we report just for a reference that the wave normal directions determined by the above mentioned data are roughly $65^\circ \pm 15^\circ$. In order to determine the wave normal direction more correctly, the wave distribution function (WDF) method introduced by Storey and Lefeuvre [1979, 1980] must be used. The wave normal direction of the He^+ mode emissions has not yet been analyzed, because they have not enough magnetic field components in general.

3. RAY PATHS CALCULATIONS OF THE ELF EMISSIONS

We have developed a three-dimensional ray tracing program including the effects of ions. The propagation characteristics of the ICW mode emissions can be examined by ray tracing. In this section we will present the results of three-dimensional ray tracing for ion mode waves both in cold and hot plasma model.

3.1. Techniques of three-dimensional ray tracing

The fundamental differential equations of ray tracing were derived by Haselgrove [1955] and Kimura [1966]. We integrate these differential equations numerically, and the propagation direction and the wave normal direction of a wave in a plasma can be determined. Our three-dimensional ray tracing program is applicable to the ion cyclotron modes in either a cold or a hot plasma.

The ray paths are represented by the spherical coordinate system (r, θ, ϕ) , where r, θ, ϕ are the geocentric distance, the geomagnetic colatitude, and the geomagnetic longitude in a dipolar spherical coordinate system, respectively. In order to express the direction of the wave normal vector k we use the coordinate system represented by (ψ, η) shown in Figure 9, where ψ is a polar angle between k and the geomagnetic field B_0 , and η is the azimuthal angle between the component of k perpendicular to B_0 and the plane which is made by B_0 and the radial direction.

As for the geomagnetic field, a dipole model is adopted. For the plasma density profile in the magnetosphere a diffusive equilibrium (DE) model [Angerami and Thomas, 1964] is used. Therefore if a ray tracing is started with an initial wave normal direction in the geomagnetic meridian plane, the ray paths do not deviate from the same geomagnetic meridian plane. We have, however, found that when the initial wave normal direction deviates from a geomagnetic meridian plane, we have found that the ray path first deviates from the initial meridian plane and is finally trapped in a different meridian plane. In the following therefore the ray tracing results are those obtained by the two-dimensional ray tracing in a geomagnetic meridian plane.

3.2. Generation mechanism of the emissions

3.2.1. *Condition of the plasmasphere.* As mentioned in section 2.2.1, the existence of H^+ and He^+ mode emissions suggests a relatively high content of He^+ and O^+ ions at altitudes as high as 10,000 km.

Figure 10 shows the electron temperature observed by the TED (temperature and energy distribution of plasma) instruments on board the Akebono satellite in the equatorial region on July 12, 1989, when the two band emissions were observed. In this figure, GLON, GLAT, and ILAT stand for geographic longitude, geographic latitude, and invariant latitude, respectively. (This figure was provided by K. Oyama and T. Abe of the Akebono TED project team.) It is remarkable that the electron temperature is about 5000°. According to the temperature observations by the Taiyo satellite during the rather low solar activity period of 1975, the electron temperature was around 2500° for $L = 1.3$ to 1.7 [Oya and Oyama, 1985]. Five thousand degrees is a very high temperature, maybe due to a high solar activity in 1989. Although it is not certain that the magnetospheric plasma is in perfect thermal equilibrium with the same temperature for electron and ions, such a high temperature is favorable to explain the existence of heavy ions such as He^+ and O^+ in high altitudes based on the DE density model. To theoretically clarify our observed results of the ELF emissions by using ray tracing, we have to know the relative concentration of He^+ and O^+ ions, and the flux of nonthermal H^+ ions. However, because the trajectories of the Akebono satellite are within the inner radiation belt, the ion mass spectrometer and low-energy particle

analyzer on board the Akebono satellite cannot measure these quantities correctly. We therefore assume the following plasma parameters in the DE model: 44% of H^+ , 8% of He^+ , 48% of O^+ , $1.5 \times 10^4 \text{ cm}^{-3}$ of the electron density at the reference altitude of 1100 km, and the particle temperature of 5000°. Since we do not have any information of the relative constituents of ions by the Akebono satellite around the geomagnetic equator, the relative constituents of ions were referred to the data observed by the ISS-b satellite [Radio Research Laboratories Ministry of Posts and Telecommunications, 1985].

3.2.2. *Dispersion curve of the emissions.* In the presence of H^+ , He^+ , and O^+ ions in the magnetospheric plasma an ω - k diagram in the frequency range around their ion cyclotron frequencies is as shown in Figure 11. In this ω - k diagram, ω is normalized by Ω_{H^+} , and k is normalized by Ω_{H^+}/V_A , where V_A is the Alfvén velocity. The solid lines and dashed lines in the figure correspond to the wave normal angle ψ of 0° and 90°, respectively. The coarsely hatched region corresponds to the region of left-handed polarization and the finely hatched region corresponds to that of right-handed polarization.

In the presence of H^+ , He^+ , and O^+ ions, there are three characteristic regions of ICW modes indicated by the symbols " H^+ ," " He^+ ," and " O^+ " and the electron mode connected to the whistler mode is indicated by the symbol " e ." There are cutoff frequencies and crossover frequencies. It is very important to note that there are two band gaps between Ω_{He^+} and Ω_{LHR2} , and between Ω_{O^+} and Ω_{LHR3} , where Ω_{LHRi} denotes the i th lower hybrid resonance (LHR) angular frequency. In this dispersion

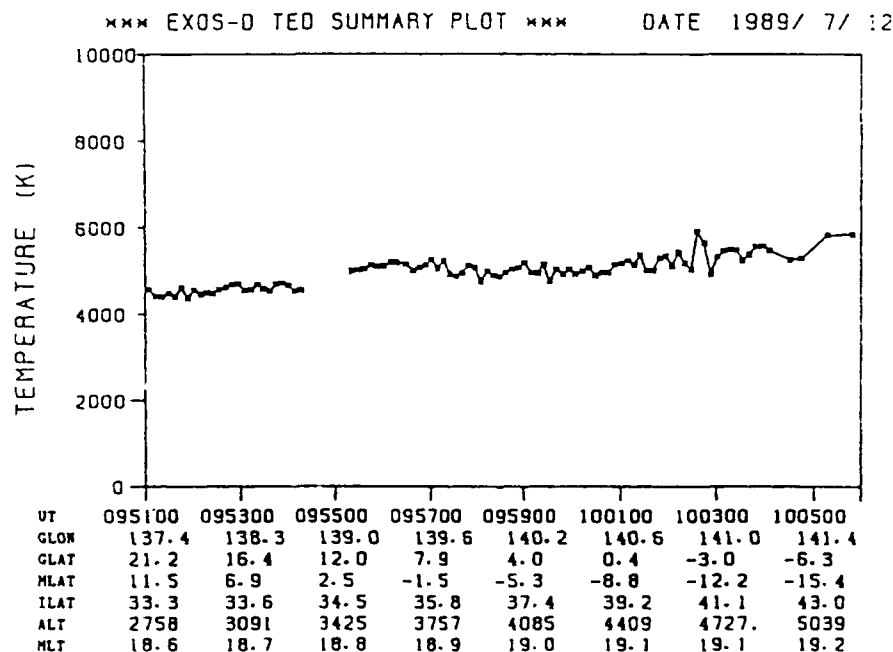


Fig. 10. Electron temperature observed by the TED instruments while both the H^+ mode and the He^+ mode emissions were observed on July 12, 1989.

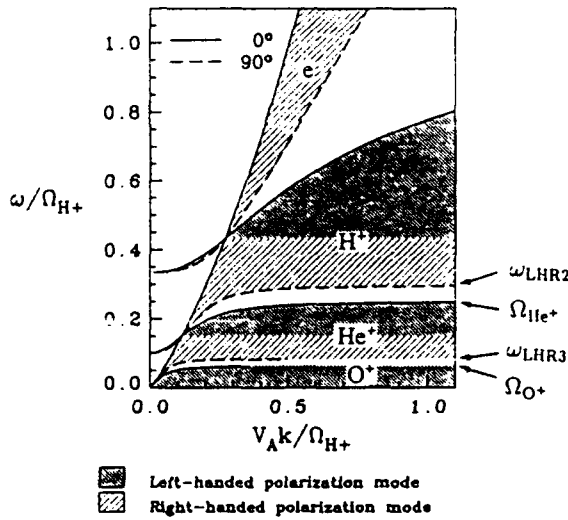


Fig. 11. Dispersion curve for the electromagnetic ion mode waves in the presence of H^+ , He^+ and O^+ ions. The relative constituents of ions there are H^+ : 85.0%, He^+ : 9.0%, and O^+ : 6.0%.

diagram the H^+ mode emissions correspond to the region indicated by the symbol " H^+ " and the He^+ mode emissions correspond to the region indicated by the symbol " He^+ ." It should be remembered that H^+ mode waves cannot exist without a sufficient content of He^+ ions, at least, more than a few percent. In the same way, He^+ mode wave cannot exist without a sufficient content of O^+ ions.

Figure 12 is an ω - k diagram which is calculated by using the relative constituents of ions; H^+ : 91.5%, He^+ : 7.1% and O^+ : 1.4%. These percentages correspond to those at an altitude of $1.5 R_e$ (R_e is the radius of the Earth, that is, $\approx 6,370$ km) on the geomagnetic equatorial plane, according to the DE density model with the parameters mentioned previously. The axes on the right and on the top indicate the real wave frequency and $|k|$ respectively at an altitude of $1.5 R_e$ on the geomagnetic equatorial plane.

In section 2.3.1 we roughly estimated the refractive indices of H^+ mode and He^+ mode by (2) and found that the refractive index of H^+ mode is from 100 to 200 and that of He^+ mode is 400 to 600 at an altitude range from 3000 to 4000 km. In the ω - k diagram a refractive index is indicated by the gradient of a line passing through the origin. Considering the refractive indices and frequency range of the observed emissions, we can presume that these emissions are located in the cross-hatched regions.

3.2.3. Growth rate of the emissions. Now we will discuss the generation mechanism of the emissions. According to Roux et al. [1982], the ICW mode waves which they observed by the GEOS 1 and 2 satellite near the plasmapause are assumed to be generated by the ion cyclotron resonant instability caused by a temperature anisotropy of hot H^+ ions around the geomagnetic equatorial plane. On the other hand, our observations have been made within the inner radiation belt, so that we can expect

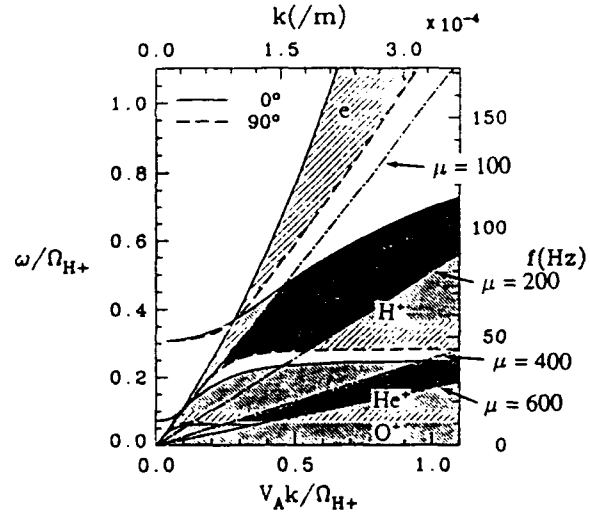


Fig. 12. Dispersion curve at an altitude of $1.5 R_e$ on the geomagnetic equator in the DE model. The relative constituents of ions there are H^+ : 91.5%, He^+ : 7.1%, and O^+ : 1.4%.

a possible existence of an abundant energy source for the ICW mode waves which we observed in the area. In this paper we assume that these ICW mode emissions are generated by an ion cyclotron resonant instability similar to the generation mechanism suggested by Roux et al. [1982].

In our calculations the relative constituents of ions at $1.5 R_e$ as mentioned before are used. Particle temperatures are assumed as follows; 10% of H^+ ions are hot with an anisotropy of $T_{hot\perp}/T_{hot\parallel} > 1$, and He^+ , O^+ and 90% of H^+ ions are background particles with an energy of 1.5 eV. As for the hot H^+ , we have calculated for $T_{hot\parallel}$ from 100 eV to 100 MeV, with two temperature anisotropy cases; $T_{hot\perp}/T_{hot\parallel} = 2$ and $T_{hot\perp}/T_{hot\parallel} = 3$.

Figure 13 shows the growth rate γ (imaginary part of ω) normalized by Ω_{H^+} of the ICW mode waves, versus the thermal energy of the hot H^+ ions ($T_{hot\parallel}$) parallel to the geomagnetic field line, calculated by using the parameters mentioned above. The wave normal angle ψ is assumed to be 0° . In this figure the growth rates of H^+ mode and He^+ mode wave are shown. However, for He^+ mode wave the growth rate is not plotted for $T_{hot\parallel}$ larger than 100 keV, because the real part ω_r and the imaginary part γ become comparable and the growth rate is meaningless. From this figure it is seen that the hot H^+ ions with energies of 100 keV are most effective for the growth of the H^+ mode waves. It is because the Alfvén velocity is of the same order as the thermal velocity of $T \approx 100$ keV.

Figure 14 shows the dependency of the growth rate of the H^+ mode waves on the wave normal angle ψ , where the thermal velocity of the hot H^+ is assumed to be 100 keV and the temperature anisotropy $T_{hot\perp}/T_{hot\parallel}$ is 2. As shown in this figure, the ICW mode waves interact most closely with the hot H^+ when the wave normal direction is parallel to the geomagnetic field line.

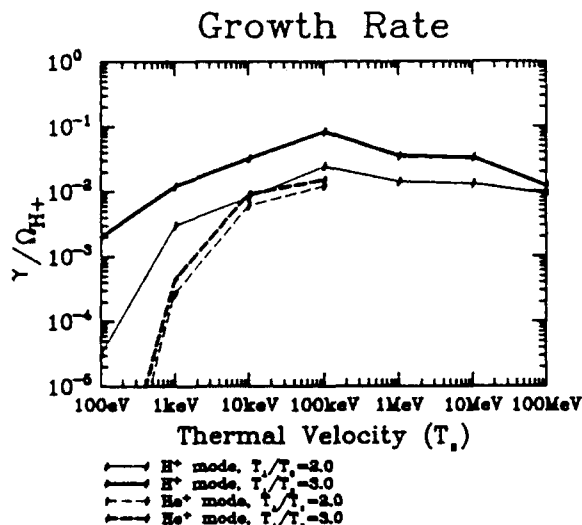


Fig. 13. Growth rate γ of the ICW mode waves caused by a temperature anisotropy of the hot H^+ ions. The horizontal axis indicates $T_{h\alpha//}$ of the hot H^+ . The angle ψ is assumed to be 0° .

3.3. Results of cold plasma ray tracing

In this section we will show the ray paths for the H^+ and He^+ mode waves calculated by ray tracing in a cold plasma model and discuss the general propagation characteristics of the emissions and the spatial distribution where they were observed.

As mentioned in section 3.2.3, we have studied that the ICW mode emissions can be generated by ion cyclotron resonant instability caused by a temperature anisotropy of hot H^+ ions around the geomagnetic equatorial region. We have also learned that most favorable wave normal angle with a geomagnetic field line for the ion cyclotron resonant instability is nearly zero. We therefore start our ray tracing at the geomagnetic equatorial region with $\psi = 0^\circ$. The plasma parameters used in the DE model are the same as those mentioned in section 3.2.2. Each ray path is initiated from the geomagnetic equatorial plane at an L value from 1.4 to 2.9 with a frequency at which the ion cyclotron resonant instability has the maximum growth rate at the start point where the thermal velocity of the hot H^+ is assumed to be 100 keV and the temperature anisotropy $T_{h\alpha\perp}/T_{h\alpha//}$ is 2.

Figure 15a shows the ray paths calculated for the H^+ mode waves, and Figure 15b shows those for the He^+ mode waves. The arrows along each ray path represent the wave normal directions. As shown in these figures, H^+ mode and He^+ mode waves propagate along the geomagnetic field lines, and are reflected back at the lower hybrid resonance condition. Once reflected, these waves propagate back and forth along almost the same ray paths nearly along a geomagnetic field line and are trapped within a limited geomagnetic latitude range. From these two figures it is evident that He^+ mode waves can propagate over a wider range in latitude than H^+ mode waves.

Figure 16 shows the related quantities calculated along a ray path when the H^+ mode wave is initiated from the geomagnetic

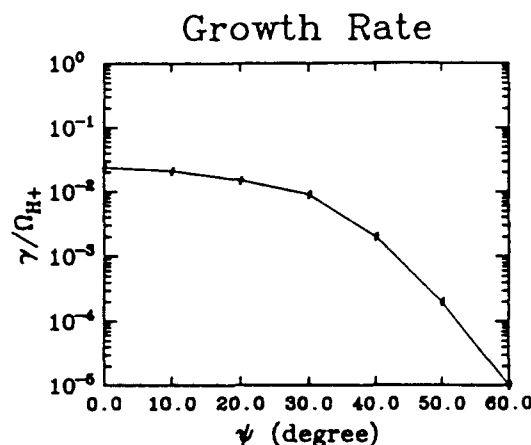


Fig. 14. Relation between the growth rate γ of the H^+ mode wave and the angle ψ , where $T_{h\alpha//} = 100$ keV and $T_{h\alpha\perp}/T_{h\alpha//} = 2$.

equator at an L value of 1.5. In the figure the symbol "o" indicates the initial point of the ray path. In Figure 16a, r indicates the radial position of the ray path and t is the propagation time. We can find that the wave bounces with periodic trajectory of about every 5.5 s. In Figure 16b, "ray" indicates the ray direction with respect to the geomagnetic field line. It is evident that the H^+ mode wave is trapped on a same geomagnetic field line. Figure 16c shows the variation of the wave number along the ray path, and Figure 16d shows the relation between r and ψ . It turns out that the wave normal angle gradually increases and turns to a direction nearly perpendicular to the geomagnetic field line. That is, the wave gradually changes to be electrostatic.

Figure 17 shows the similar quantities along the ray path of He^+ mode started at an L value of 1.5. As shown in Figure 17a, the He^+ mode wave stays as long as about 15 s around the geomagnetic equatorial region and bounces back from the reflecting point very quickly, the interval of bouncing back to the geomagnetic equatorial plane is about 36 s. The other propagation characteristics of the He^+ mode wave are similar to those of H^+ mode wave, that is, the wave propagates along the same geomagnetic field line, the wave normal gradually increases and approaches perpendicular to the geomagnetic field line and the wave gradually changes to be electrostatic.

Thus the ray tracing in a cold plasma model has clarified that the latitudinal range of the observed emissions. The azimuthal direction of the wave normal represented by η was not mentioned in the above discussion, because the wave normal direction stays in the same meridian plane (i.e., $\eta = 0^\circ$) in the dipole magnetic field model, due to the assumption that the initial wave normal is directed along the geomagnetic field line. It has been found that even if a ray tracing is started with an initial wave normal direction off from the geomagnetic meridian plane the ray paths finally follow nearly along one meridian plane, though such ray paths are not shown here. To determine the three-dimensional effect, we have to use the international geomagnetic reference field model (IGRF) for ray tracing.

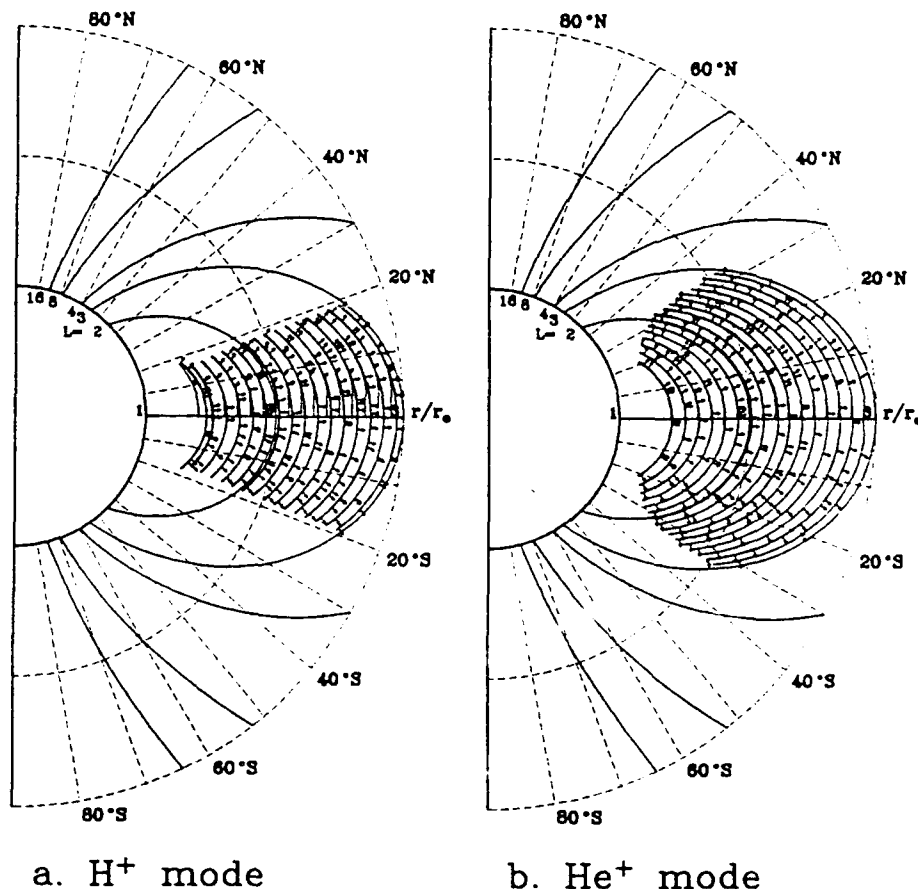


Fig. 15. Ray paths of the ICW calculated by ray tracing method in cold plasma model. Each ray path is initiated from the geomagnetic equatorial plane at an L value from 1.4 to 2.9 with a frequency of maximum growth rate at the start point where $T_{\text{hot}\perp} = 100$ keV and $T_{\text{hot}\perp}/T_{\text{hot}\parallel} = 2$.

3.4. Results of hot plasma ray tracing

In this section we will show the ray paths calculated in a hot plasma. For ray tracing in a hot plasma we have extended our hot plasma ray tracing program [Hashimoto et al., 1987] to the case where the effects of ions are included. By this ray tracing we can calculate the total growth and damping along the ray paths.

In our hot plasma ray tracing the angular frequency of the wave ω is treated as a real number, and the wave vector k is regarded as a complex number, that is, $k = k_r + jk_i$. Then, the total integrated growth or damping along the ray path is calculated by the following integrations:

$$\Gamma = - \int_{\text{path}} k_i \cdot dr = - \int_{\text{path}} k_i \cdot v_g dt \quad (4)$$

In the actual ray tracing we have imposed a condition that the ratio of $|k_i|$ to $|k_r|$ is required to be less than 0.1, and the ray

tracing was terminated when this condition is not satisfied. It is because for a relatively large $|k_i|$ condition the ray path obtained by ray tracing becomes meaningless.

Figure 18 shows the results of the ray tracing for the H⁺ mode wave, which is initiated from the geomagnetic equator at an L value of 1.5 with its wave normal direction along the geomagnetic field line, using the plasma model with 10% of hot H⁺ ions, whose thermal velocity is assumed to be 100 keV and the temperature anisotropy $T_{\text{hot}\perp}/T_{\text{hot}\parallel}$ is 2, all along the ray paths. Figures 18a-18c show the ray path, the variation of the wave normal angle ψ , and the total integrated growth Γ along the ray path, respectively. In the figures the thick lines indicate the parts where the wave is growing.

In general, the propagation characteristics of the wave is very similar to those shown by the cold plasma ray tracing method. As shown in Figure 18b, since the angle ψ gradually turns to a direction perpendicular to the geomagnetic field line, the wave grows at first and then stops growing. In this case the wave

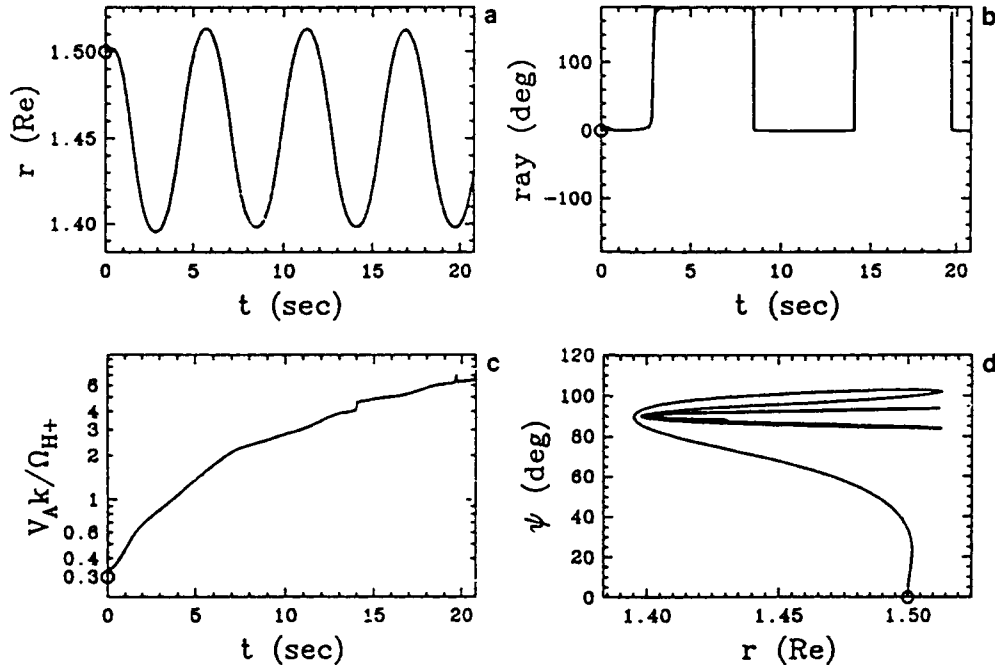


Fig. 16. Parameters of the H^+ mode wave along the ray path. The wave is initiated from the geomagnetic equator at an L value of 1.5 with its wave normal parallel to the geomagnetic field with a frequency of maximum growth rate at the start point where $T_{\text{hot}\parallel} = 100$ keV and $T_{\text{hot}\perp}/T_{\text{hot}\parallel} = 2$.

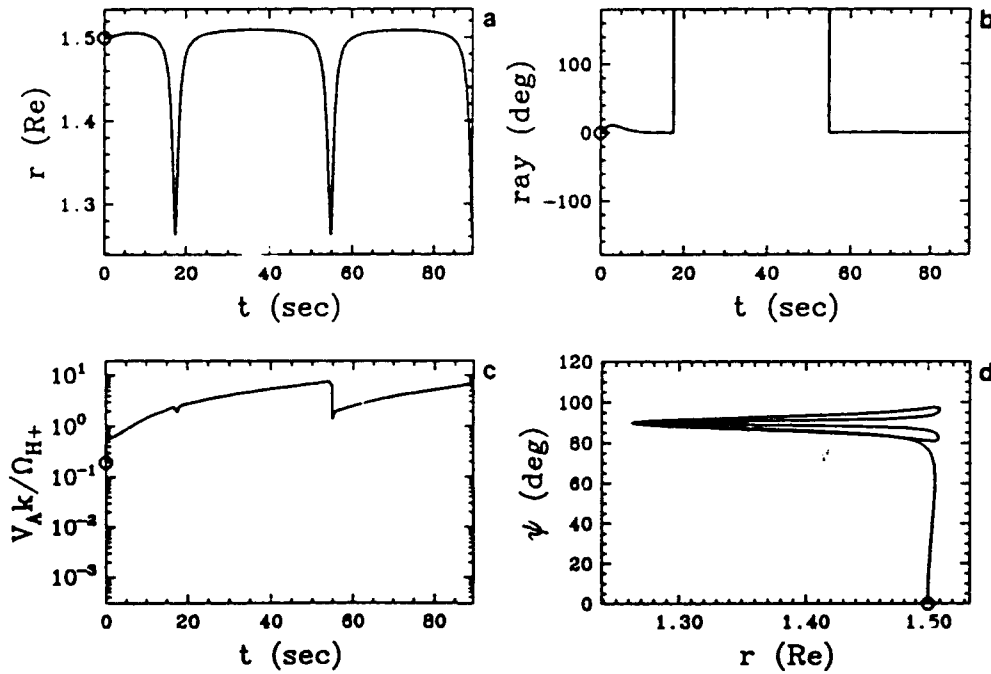


Fig. 17. Parameters of the He^+ mode wave along the ray path. The wave is initiated from the geomagnetic equator at an L value of 1.5 with its wave normal parallel to the geomagnetic field with a frequency of maximum growth rate at the start point where $T_{\text{hot}\parallel} = 100$ keV and $T_{\text{hot}\perp}/T_{\text{hot}\parallel} = 2$.

grows up to nearly 30 dB in two bounces. The region where the wave grows is within the geomagnetic latitude of $\pm 10^\circ$.

We have tried to do the ray tracing of the H^+ mode wave by starting from various geomagnetic latitude at an L value of 1.5 with its wave normal direction along the geomagnetic field. Figure 19 shows the results that the wave can grow over most

part of the ray path. In this case the wave is initiated from the geomagnetic latitude of 5° , and the wave grows up to 36 dB. After many ray tracing analyses we have found that the H^+ mode waves which are initiated from the geomagnetic latitude range within 10° can efficiently grow and propagate in the magnetosphere.

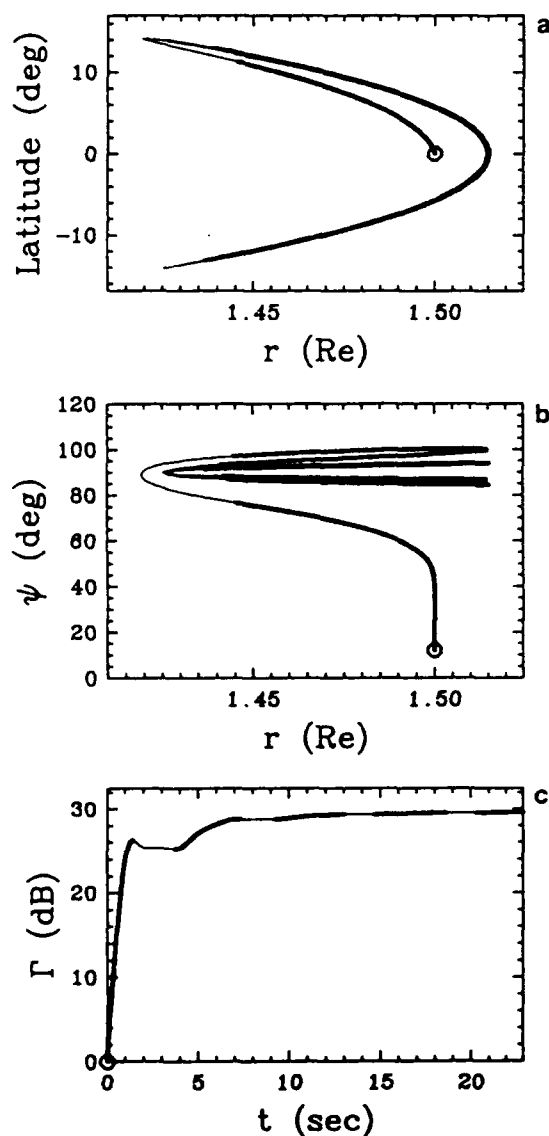


Fig. 18. Ray path and parameters of the H^+ mode wave along the ray path. The wave is initiated from the geomagnetic equator at an L value of 1.5 with its wave normal parallel to the geomagnetic field with a frequency of maximum growth rate at the start point where $T_{hot//} = 100$ keV and $T_{hot\perp}/T_{hot//} = 2$.

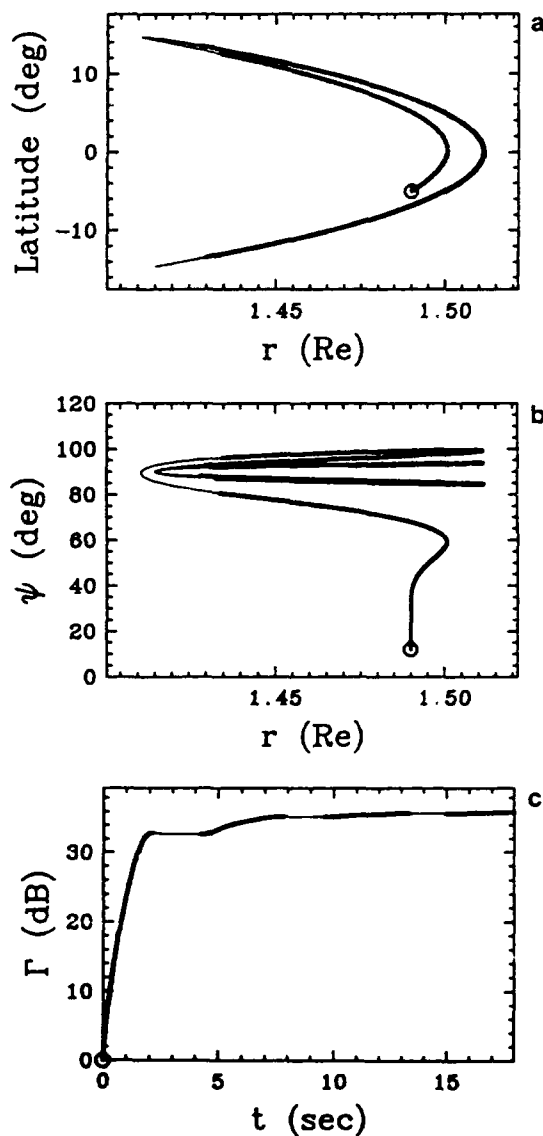


Fig. 19. Ray path and parameters of the H^+ mode wave along the ray path. The wave is initiated from the geomagnetic latitude of 5° at an L value of 1.5 with its wave normal parallel to the geomagnetic field with a frequency of maximum growth rate at the start point where $T_{hot//} = 100$ keV and $T_{hot\perp}/T_{hot//} = 2$.

On the other hand, the ray path of the He^+ mode wave is shown in Figure 20. The ray path is also initiated from the geomagnetic equator at an L value of 1.5. In comparison with the results of the H^+ mode wave, this wave does not grow so much (up to 11.5 dB at most), and is damped quickly after a couple of bouncing. These results indicate that the He^+ mode waves are less favorable to be generated than the H^+ mode waves.

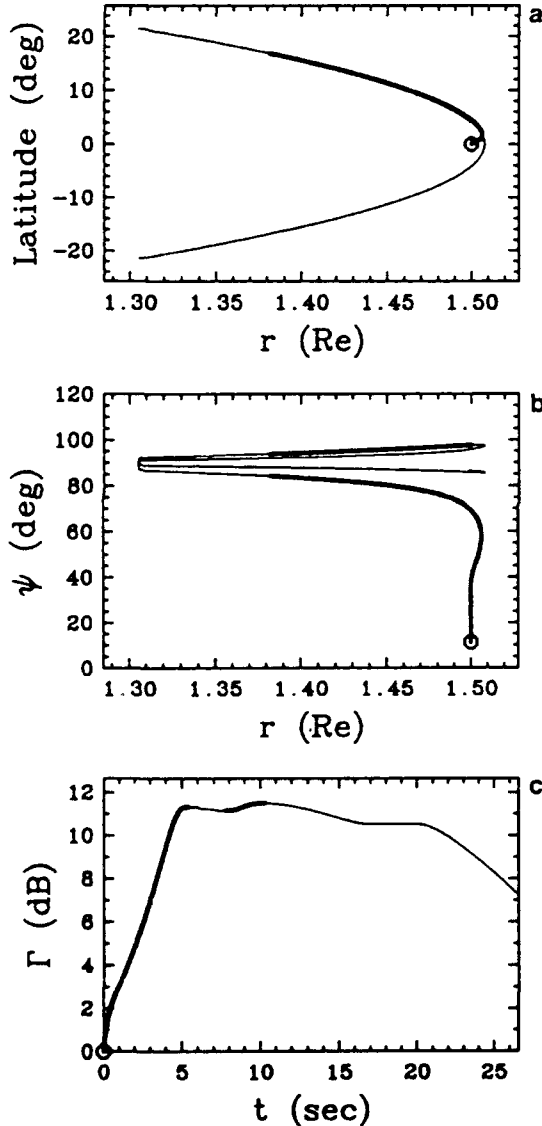


Fig. 20. Ray path and parameters of the He^+ mode wave along the ray path. The wave is initiated from the geomagnetic equator at an L value of 1.5 with its wave normal parallel to the geomagnetic field with a frequency of maximum growth rate at the start point where $T_{\text{hot}\perp} = 100$ keV and $T_{\text{hot}\perp}/T_{\text{hot}\parallel} = 2$.

4. DISCUSSIONS AND CONCLUSIONS

Our ELF observations by the Akebono satellite have revealed that bands of electromagnetic emissions which have a frequency dependence on the local cyclotron frequencies of ions are often observed in the vicinity of the geomagnetic equatorial region in a wide range of altitudes. They are identified as electromagnetic ion cyclotron waves (ICW) which might have been generated in the geomagnetic equatorial plane.

In particular, the existence of the lower band emissions, which we call the He^+ mode, requires an existence of rather high concentration of O^+ ions. This fact may be due to the low altitude of the Akebono satellite. In any case, over the region where the He^+ mode emissions are observed the concentration of O^+ ions should be high.

The high electron temperature (4000 ~ 5000 K) observed by the TED instruments in 1989 may be due to a high solar activity. Although it is not certain that the magnetospheric plasma is in perfect thermal equilibrium with an equal temperature for electrons and ions, such a high temperature is favorable to explain the existence of the ICW mode emissions. However, the relation between solar activity and occurrence probability of the ICW mode wave is not clear at present, because we began to observe only 2 years ago. In the same way, as for the statistical characteristics of the emissions, they will become clear in 1 or 2 years as we analyze more wave data.

The ray tracing method for ion mode waves has been applied to the observed emissions for both cold and hot plasma models. We have found that these ICW mode waves propagate along the geomagnetic field lines, and are trapped within a limited geomagnetic latitude range with multiple bounces by the LHR reflection. In the process of multiple bounces the wave gradually turns out to be electrostatic with its wave normal increasing and turning nearly perpendicular to the geomagnetic field line.

The total integrated growth Γ along the ray path is estimated with our hot plasma ray tracing method, on the assumption that the ion cyclotron resonant instability due to the temperature anisotropy of the hot H^+ ions causes the generation of the ICW mode wave. We have found that these waves can grow sufficiently when they are initiated around the geomagnetic equator with its wave normal vector parallel to the geomagnetic field line. Since the trajectories of the Akebono satellite around the geomagnetic equator is within the inner radiation belt, an abundant energy source for the ion cyclotron resonant instability may possibly exist, although the ion mass spectrometer and low-energy particle analyzer could not have measured these quantities correctly. According to the PWS (plasma waves and sounder experiments) measurement for frequencies from 20 kHz to 5 MHz, upper hybrid resonance (UHR) emissions are sharply intensified strictly at the magnetic equator [Oya et al., 1990]. This fact implies that an energy source is always present there.

In conclusion, the generation mechanism of these emissions is considered as follows. That is, the source region of the H^+ and He^+ mode emissions is within the geomagnetic latitude of 10° and the energy source is a temperature anisotropy of the hot H^+ which exists within the inner radiation belt. In that condition

these emissions can be generated and grow while bouncing back and forth, and then they gradually turn out to be electrostatic and are damped. In this way the back and forth waves can coexist around the geomagnetic equatorial region.

Acknowledgments. We would like to thank other EXOS-D VLF team members for their cooperation in developing the spacecraft borne VLF instrument and in data analyses. We also thank K. Oyama and T. Abe for their helpful discussion and for providing electron temperature data observed by the Akebono TED. We also appreciate helpful comments and discussion given by H. Matsumoto and Y. Omura of RASC, Kyoto University. This project was financially supported by Grant-in-Aid for Scientific Research project 63420016, Ministry of Education, Science and Culture of Japan.

REFERENCES

- Angerami, J. J., and J. O. Thomas, The distribution of ions and electrons in the Earth's exosphere, *J. Geophys. Res.*, **69**, 4537-4560, 1964.
- Gendrin, R., M. Ashour-Abdalla, Y. Omura, and K. Quest, Linear analysis of ion cyclotron interaction in a multicomponent plasma, *J. Geophys. Res.*, **89**, 9119-9124, 1984.
- Gomberoff, L., and P. Vega, Effect of He^+ temperature and thermal anisotropy on the electromagnetic ion cyclotron instability in the geosynchronous region of GEOS 2, *J. Geophys. Res.*, **92**, 7728-7732, 1987.
- Haselgrove, J., Ray theory and a new method for ray tracing, in *Proceedings of the Cambridge Conference on Physics of the Ionosphere*, p. 355, Physical Society, London, 1955.
- Hashimoto, K., K. Yamaashi, and I. Kimura, Three-dimensional ray tracing of electrostatic cyclotron harmonic waves and Z mode electromagnetic waves in the magnetosphere, *Radio Sci.*, **22**, 579-594, 1987.
- Kimura, I., Effects of ions on whistler-mode ray tracing, *Radio Sci.*, **1**, 269-283, 1966.
- Kimura, I., K. Hashimoto, I. Nagano, T. Okada, M. Yamamoto, T. Yoshino, H. Matsumoto, M. Ejiri, and K. Hayashi, VLF observations by the Akebono (EXOS-D) satellite, *J. Geomagn. Geoelectr.*, **42**, 459-478, 1990.
- Mauk, B. H., C. E. McIlwain, and R. L. McPherron, Helium cyclotron resonance within the Earth's magnetosphere, *Geophys. Res. Lett.*, **8**, 103-106, 1981.
- Omura, Y., M. Ashour-Abdalla, R. Gendrin, and K. Quest, Heating of thermal helium in the equatorial magnetosphere: A simulation study, *J. Geophys. Res.*, **90**, 8281-8292, 1985.
- Oya, H., and K. Oyama, Data from TAIYO satellite, report, Inst. of Space and Astronaut. Sci., Kanagawa, 1985.
- Oya, H., A. Morioka, K. Kobayashi, M. Iizima, T. Ono, H. Miyaoka, T. Okada, and T. Obara, Plasma Wave Observation and Sounder Experiments (PWS) using the Akebono (EXOS-D) satellite - Instrumentation and initial results including discovery of the high altitude equatorial plasma turbulence, *J. Geomagn. Geoelectr.*, **42**, 411-442, 1990.
- Radio Research Laboratories Ministry of Posts and Telecommunications, Summary plots of ionospheric parameters obtained from Ionosphere Sounding Satellite-b, report, vol.4, Tokyo, 1985.
- Rauch, J. L., and A. Roux, Ray tracing of ULF waves in a multicomponent magnetospheric plasma: Consequences for the generation mechanism of ion cyclotron waves, *J. Geophys. Res.*, **87**, 8191-8198, 1982.
- Roux, A., S. Perraut, J. L. Rauch, C. de Villedary, G. Kremser, A. Korth, and D. T. Young, Wave-particle interactions near Ω_{He^+} observed on-board GEOS 1 and 2, 2, Generation of ion cyclotron waves and heating of He^+ ions, *J. Geophys. Res.*, **87**, 8174-8190, 1982.
- Samson, J. C., and J. V. Olson, Some comments on the descriptions of the polarization states of waves, *Geophys. J. R. Astron. Soc.*, **61**, 115-129, 1980.
- Storey, L. R. O., and F. Lefeuvre, The analysis of 6-component measurements of a random electromagnetic wave field in a magnetoplasma, - 1, The direct problem, *Geophys. J. R. Astron. Soc.*, **56**, 255-269, 1979.
- Storey L. R. O., and F. Lefeuvre, The analysis of 6-component measurements of a random electromagnetic wave field in a magnetoplasma, - 2, The integration kernels, *Geophys. J. R. Astron. Soc.*, **62**, 173-194, 1980.
- Young, D. T., S. Perraut, A. Roux, C. de Villedary, R. Gendrin, A. Korth, G. Kremser, and D. Jones, Wave-particle interactions near Ω_{He^+} observed on GEOS 1 and 2, 1, Propagation of ion cyclotron waves in He^+ -rich plasma, *J. Geophys. Res.*, **86**, 6755-6772, 1981.

K. Hayashi and S. Kokubun, Institute of Geophysics, University of Tokyo, Tokyo 113, Japan.

Y. Kasahara, I. Kimura, A. Sawada, and M. Yamamoto, Department of Electrical Engineering, Kyoto University, Kyoto 606-01, Japan.

List of Reviewers

Radio Science is pleased to acknowledge the assistance of the following people who reviewed papers submitted to the special section on Radio Propagation in the Ionosphere and Magnetosphere: Theory and Application.

Alpert, Y. L.
Argo, P.
Basu, S.
Bell, T. F.
Benson, R. F.
Berkey, T.
Bernhardt, P. A.
Bullough, K.
Davies, K.
Dong, B.
Franke, S. J.
Fremouw, E. J.
Gamache, R. R.
Hamahata, H.

Hashimoto, M.
Hayakawa, M.
Hunsucker, R. D.
Inan, U. S.
Kelly, F.
Kiang, Y-W.
Kimura, I.
Knepp, D. L.
Kuo, S.
Larsen, M. F.
Lee, M.-C.
Lundborg, B.
MacDougall, J. W.
Nagano, I.

Ogura, H.
Okuzawa, T.
Ondoh, T.
Sahr, J. D.
Sales, G. S.
Shimakura, S.
Strangeways, H. J.
Suchy, K.
Wagen, J.-F.
Wernik, A. W.
Yang, C. C.
Yeh, K. C.
Ziolkowski, R. W.